EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

December 15, 2000

PROVISIONAL GUIDANCE
ON THE IMPLEMENTATION OF THE 1997 STANDARDS
FOR FEDERAL DATA ON RACE AND ETHNICITY

NOTE FOR READERS

As a follow-on to OMB's October 1997 announcement of revised government-wide standards for Federal data on race and ethnicity, the Tabulation Working Group of the Interagency Committee for the Review of Standards for Data on Race and Ethnicity has recently issued "Provisional Guidance on the Implementation of the 1997 Standards for the Collection of Federal Data on Race and Ethnicity." The guidance presented in this document is intended for any Federal agencies or organizational units that maintain, collect, or present data on race and ethnicity for Federal statistical purposes, program administrative reporting, or civil rights compliance reporting.

This is a substantially updated version of the earlier guidance that was made available in February 1999. It reflects public comments on the previous version as well as the Tabulation Working Group's further research and deliberations. The guidance, which was requested by Federal agencies and the many users of data on race and ethnicity, continues to be developed with the involvement of these constituencies. By design, this guidance does not cover all of the specific issues individual agencies will need to address during their implementation of the 1997 standards.

The guidance for implementing the 1997 standards focuses on three areas: collecting data, tabulating data, and building bridges to compare data collected under the 1997 and the 1977 standards. In some areas work is ongoing, and the guidance will be updated as additional research and analyses are completed. We expect that the guidance will evolve further as data from Census 2000 and other data collections employing the 1997 collection standards become available, as agencies address implementation issues in their respective programs, and as additional research needs are identified and addressed.

In keeping with the process that guided the development of the 1997 standards for data on race and ethnicity, we are looking forward to a continuing dialogue on this provisional guidance. We welcome your questions, comments, and suggestions.

Katherine K. Wallman
Chief Statistician

PROVISIONAL GUIDANCE

ON THE

IMPLEMENTATION

OF THE 1997 STANDARDS FOR

FEDERAL DATA ON RACE AND ETHNICITY

Prepared By

Tabulation Working Group
Interagency Committee for the Review of Standards for
Data on Race and Ethnicity

December 15, 2000

Table of Contents

PROVISIONAL GUIDANCE ON THE IMPLEMENTATION OF
THE 1997 STANDARDS FOR FEDERAL DATA ON RACE AND ETHNICITY

Prepared by
Tabulation Working Group
Interagency Committee for the Review of Standards for
Data on Race and Ethnicity

The guidance presented in this document is intended for any Federal agencies or organizational units that maintain, collect, or present data on race and ethnicity for Federal statistical purposes, program administrative reporting, or civil rights compliance reporting. The guidance complements the Federal Government's decision in October 1997 to provide an opportunity for individuals to select one or more races when responding to agency requests for data on race and ethnicity. To foster comparability across data collections carried out by various agencies, it is useful for those agencies to report responses of more than one race using some standardized tabulations or formats.

The report briefly explains why the tabulation guidelines are needed, reviews the general guidance issued when the standards were adopted in October 1997, and provides information on the criteria used in developing the guidelines. This report also addresses a larger set of implementation questions that have emerged during the working group's deliberations. Thus, the report considers:

- C   Collecting data on race and ethnicity using the 1997 standards;
- C   Tabulating Census 2000 data as well as data on race and ethnicity collected in surveys and from administrative records;
- C   Using data on race and ethnicity in applications such as legislative redistricting, civil rights monitoring and enforcement, and population estimates; and
- C   Comparing data under the 1997 and the 1977 standards when conducting trend analyses.

In addition, the appendices to the report provide the full text of reports on the research that has been conducted in two areas: approaches for collecting data on race and ethnicity, and approaches for bridging between data collected under the 1997 standards and data collected under the 1977 standards.

The guidelines are necessarily provisional pending the availability of data from Census 2000 and other data systems as the 1997 standards are implemented. The guidelines provide a general framework and are not intended to cover all aspects of problems that agencies will encounter during their implementation of the 1997 standards. In some instances, for example, specific implementation issues are being address through OMB's paperwork review of data collections. The guidelines are likely to be reviewed and refined as Federal agencies and others gain experience with data collected under the 1997 standards.

CHAPTER 1
BACKGROUND

This chapter discusses why guidance is needed for tabulating data collected using the 1997 standards, reiterates the general guidance issued when the standards were adopted in October 1997, provides clarification of several aspects of the standards, and presents the criteria that were developed for evaluating bridging methods and presenting data.

A. Need for Tabulation Guidelines

On October 30, 1997, the Office of Management and Budget (OMB) published "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity" (*Federal Register*, 62 FR 58781 - 58790)(see Appendix A). The 1997 standards reflect a change in data collection policy, making it possible for Federal agencies to collect information that reflects the increasing diversity of our Nation's population stemming from growth in interracial marriages and immigration. Under the new policy, agencies are now required to offer respondents the option of selecting one or more of the following five racial categories included in the 1997 standards:

-- **American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

-- **Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

-- **Black or African American.** A person having origins in any of the black racial groups of Africa. Terms such as "Haitian" or "Negro" can be used in addition to "Black or African American."

-- **Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

-- **White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

These five categories are the minimum set for data on race for Federal statistics, program administrative reporting, and civil rights compliance reporting.

With respect to ethnicity, the standards provide for the collection of data on whether or not a person is of "Hispanic or Latino" culture or origin. (The standards do not permit a multiple response that would indicate an ethnic heritage that is both "Hispanic or Latino" and "Not Hispanic or Latino.") This category is defined as follows:

-- **Hispanic or Latino.** A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. The term, "Spanish origin," can be used in addition to "Hispanic or Latino."

As a result of the change in policy for collecting data on race, the reporting categories used to present these data must similarly reflect this change. In keeping with the spirit of the 1997 standards, agencies cannot collect multiple responses and then report and publish data using only the five single race categories. Agencies are expected to provide as much detail as possible on the multiple race responses, consistent with agency confidentiality and data quality criteria. As provided by the standards, OMB will consider any agency variances to this policy on a case by case basis.

Based on agency research prior to the issuance of the 1997 standards, it was estimated that less than two percent of the Nation's total population was likely to identify with more than one race. This percentage may increase as those who identify with more than one racial heritage become aware of the opportunity to report more than one race. As compared with data collected on adults, data collected on children and youth, however, are likely to reflect larger numbers and percentages of respondents reporting themselves as belonging to more than one racial group.

In the early years of the standards' implementation, there will be issues of data quality and confidentiality related to sample size that may restrict the amount of data that can be published for some combinations of multiple race responses. Over time, however, the size of these data cells may increase. It should be noted that such data quality and confidentiality problems for small population groups also existed under the 1977 standards, where sample sizes sometimes prevented presentation of data on certain population groups such as American Indians. The possible multiple race combinations under the 1997 standards, some with small data cells, serve to make such data quality concerns more apparent. Some balance will need to be struck between having a tabulation showing the full distribution of all possible combinations of multiple race responses and presenting only the minimum -- that is, a single aggregate of people who reported more than one race.

When the standards were announced on October 30, 1997, they became effective immediately for all new and revised Federal record keeping or reporting requirements that included data on race and ethnicity. All existing Federal record keeping or reporting requirements must be made consistent with the provisions of the 1997 standards at the time they are submitted to OMB for extension through the reports clearance process, or no later than January 1, 2003. As provided by the standards, an agency must make a request to OMB for any variation from the standards.

B. General Guidelines for Tabulating Data on Race

In response to concerns that had been raised about how Federal agencies would tabulate multiple race responses, OMB in the October 30, 1997, *Federal Register* Notice issued the following general guidance:

C   Consistent with criteria for confidentiality and data quality, the tabulation procedures used by the agencies should result in the production of as much detailed information on race and ethnicity as possible.

C   Guidelines for tabulation ultimately must meet the needs of at least two groups within the Federal Government, with the overriding objective of providing the most accurate and informative body of data. The first group is composed of those Federal Government officials charged with carrying out constitutional and legislative mandates, such as redistricting legislatures, enforcing civil rights laws, and monitoring progress in anti-discrimination programs. (The legislative redistricting file produced by the Bureau of the Census, also known as the Public Law 94-171 file, is an example of a file meeting such legislative needs.) The second group consists of the staff of Federal statistical agencies producing and analyzing data that are used to monitor economic and social conditions and trends.

C   Many of the needs of the first group can be met with an initial tabulation that provides, consistent with standards for data quality and confidentiality, the full detail of racial reporting; that is, the number of people reporting in each single race category and the number reporting in each of the possible combinations of races, which would add to the total population.

C   Depending on the judgment of users, the combinations of multiple responses could be collapsed.

   (1) One method would be to provide separate totals for those reporting in the most common multiple race combinations and to collapse the data for other less frequently reported combinations. The specifics of the collapsed distributions would be dependent on the results of particular data collections.

   (2) A second method would be to report the total selecting each particular race, whether alone or in combination with other races. These totals would represent upper bounds on the size of the populations who identified with each of the racial categories. In some cases, this latter method could be used for comparing data collected under the old standards with data collected under the 1997 standards.

C   It is important that Federal agencies with the same or closely related responsibilities adopt consistent tabulation methods.

C   Regardless of the method chosen for collapsing multiple race responses, Federal agencies must make available the total number reporting more than one race, if confidentiality and data quality requirements can be met, in order to ensure that any changes in response patterns resulting from the 1997 standards can be monitored over time.

- Different tabulation procedures might be required to meet various needs of Federal agencies for data on race. Nevertheless, Federal agencies often need to compare racial and ethnic data. Hence, some standardization of tabulation categories for reporting data on race is desirable to facilitate such comparisons.

The October 30, 1997, *Federal Register* Notice identified four areas where further research was needed on how to tabulate data:

(1) How should the data be used to evaluate conformance with program objectives in the area of equal employment opportunity and other anti-discrimination programs?

(2) How should the decennial census data for many small population groups with multiple racial heritages be used to develop sample designs and survey controls for major demographic surveys?

(3) How should the 1997 standards be introduced in the vital statistics program which obtains the number of births and deaths from administrative records, but uses intercensal population estimates in determining the rates of births and deaths?

(4) And more generally, how can meaningful comparisons be made of data collected under the previous standards and data that will be collected under the 1997 standards?

In order to address these and other issues and to ensure that tabulation methodologies would be carefully developed and coordinated among the Federal agencies, OMB assembled a group of statistical and policy analysts drawn from the Federal agencies that generate or use these data. This group has considered tabulation issues and developed the provisional guidance that is presented in this report for use by Federal agencies. The work of this group has included: (1) a review of Federal data needs and uses to ensure that the tabulation guidelines produce data that meet statutory and program requirements; (2) cognitive testing of the wording of questions; (3) cognitive testing of a form for reporting aggregate data; (4) evaluation of different methods of bridging from the 1997 to the 1977 standards; and (5) development of guidelines for presenting data on multiple race responses that meet accepted data quality and confidentiality standards.

The tabulation guidance in this report is necessarily provisional pending the availability of Census 2000 data and other data series as the 1997 standards are implemented. These guidelines will be reviewed and modified as agencies and other data users gain experience with data collected using the 1997 standards.

C. Interpretation of Self-Reported Data on Race

It is important to remember that the Federal racial and ethnic data categories are social-political constructs and that they should not be interpreted as being genetic, biological, or anthropological in nature. Data on race and ethnicity have historically been collected in the decennial census,

but the categories for collecting and tabulating these data have changed numerous times. These changes have reflected the shifts in the racial makeup of the population and changes in social attitudes and political concerns. The standard was developed in the mid-1970's in large measure to provide comparable data to monitor equal access in areas such as housing, education, mortgage lending, health care services, and employment for population groups that historically had experienced discrimination and differential treatment because of their race or ethnicity. By using the standard to tabulate data in these areas by race and ethnicity, it is possible to compare disparities across data systems. While the Federal categories provide a standardized format for purposes of collecting and presenting data on race and ethnicity, the standard was not designed to capture the full complexity of race and ethnicity in the United States. This context is important for understanding why the Federal Government collects data on race and ethnicity and for interpreting these data.

The 1997 standards emphasize self-reporting or self-identification as the preferred method for collecting data on race and ethnicity. The standards do not establish criteria or qualifications (such as blood quantum levels) that are to be used in determining a particular individual's racial or ethnic classification. They do not tell an individual who he or she is, or specify how an individual should classify himself or herself. Self-identification for race and Hispanic or Latino origin means that the responses are based on self-perception and therefore are subjective, but by definition, the responses are accurate. In situations where self-reporting is not practicable or feasible, such as identification by personnel of funeral homes, observer identification may be used. Because the 1997 standard allows individuals to report one or more races, the importance of self-identification is underscored; it is generally difficult for observers to report an individual's multiple racial heritages.

As mentioned above, a consequence of using self-identification is that unless a person is purposely misreporting, there are no wrong answers even if "objective" clues suggest otherwise. This contrasts with the collection of information on other demographic characteristics. For example, if someone born on January 1, 1950, indicates that he or she is 30 years of age when asked on January 1, 2000, the researcher views this as an error and corrects the information. The use of self-identification coupled with the social nature of race also results in situations where an individual's response to questions on race may change over time as a result of the maturation process, the particular situation, and the changing environment. If a data collection strategy used to measure age were to produce such results, that strategy would be abandoned; in the case of race, however, such change is acceptable and expected. The dynamic nature of the concept of race and how the population views it as well as the circumstances under which the data are collected need to be considered in the interpretation and analysis of these data.

An example of how the social dimension of race is incorporated into the collection of information is the large increase in the percent of the population identifying as American Indians between the 1970 and 1980 censuses. A standard interpretation of population increase focuses on the basic demographic processes of mortality, natality, and immigration. In this instance, these processes could not explain this increase. On the other hand, it would be equally wrong to

conclude that persons misreported their race in either 1970 or 1980. The increase reflects societal changes related to the perception of American Indian heritage and how these changes affect how individuals self-identify.

Another example relates to multiple race identity. One way to define "objectively" a multiple race individual would be to assess the race of each of the parents. It is possible to identify a population in this way from data sets that collect information on the characteristics of family members. This approach becomes increasingly complex in situations where parents are also of multiple races, leading to questions about how much of a given racial heritage the respondent considers "enough" to report. Analyses show that not all persons with parents of different races identify as having more than one race. It also appears that the probability of identifying multiple races changes with the age of the individual as well as with the specific races of the parents. How individuals come to their self-identification is an important research issue that would require collection of data beyond the minimum set of categories and which would not be feasible to incorporate into all data collection systems.

In interpreting and evaluating the results of analyses that are based on data collected using the 1997 standards, it is important to remember that the approach can only capture selected aspects of a complex dimension, and that what is captured will be affected by strong and complex social processes. As more detailed data are collected, and analyses that directly address these social processes are conducted, the interpretation of the information collected more routinely under the 1997 standard will be clearer.

D. Points of Clarification Regarding the 1997 Standards

This section elaborates on several points in the standards that have been a source of confusion for some users.

Under the 1997 standards, "Hispanic or Latino" is clearly designated as an *ethnicity* and not as a race. Whether or not an individual is Hispanic or Latino, every effort should be made to ascertain the race or races with which an individual identifies.

The two-question format, with the ethnicity question preceding the race question, should be used when information is collected through self-identification. Although the standards permit the use of a combined question when collecting data by observer identification, **the use of the two-question format is strongly encouraged even where observer identification is used**. Regardless of the question format, observers are expected to attempt to identify the individual's race(s).

The standards require that at a minimum the total number of persons identifying with more than one race be reported. (A response that includes, for example, two or more Asian groups is not a multiple race response.) It is stressed that this is a **minimum**; agencies are strongly encouraged to report detailed information on specific racial combinations subject to constraints of data

reliability and confidentiality standards. In this regard, agencies are expected to report only those categories that meet their current reliability and confidentiality standards. Thus, the reporting of individual categories is likely to be more detailed when the overall racial distribution is reported than when characteristics by race (such as, for example, income by race) are reported.

The following wording in the standards concerning the reporting of data when the combined question is used is clarified in the paragraph below:

> "In cases where data on multiple responses are collapsed, the total number of respondents reporting 'Hispanic or Latino and one or more races' and the total number of respondents reporting 'more than one race' (regardless of ethnicity) shall be provided." (Section 2b of the standards)

A complete tabulation of race by ethnicity should always be reported when confidentiality permits. If not, at least ethnicity by the single races and ethnicity for those reporting more than one race should be given. Thus, an Hispanic or Latino respondent reporting one race should be reported both as Hispanic or Latino and as a member of that single race. Reporting a composite - - such as the number of people who responded "Hispanic or Latino" *and* "more than one race" -- is a **minimum** that should be used only if more detailed reporting would violate data reliability and confidentiality standards.

The rules discussed in Section 4 of the 1997 standards concerning the presentation of data on race and ethnicity **under special circumstances** are not to be invoked unilaterally by an agency. If the agency believes the standard categories are inappropriate, the agency must request a specific variance from OMB.

The 1997 standards **do not** include an "other race" category. For Census 2000, OMB granted an exception to the Census Bureau to use a category called "Some Other Race." OMB has also granted an exception to the National Center for Health Statistics to include "Some Other Race" on the U.S. standard birth and death certificates in order to maintain comparability between the demographically related data systems of vital statistics and the decennial census.

E. Criteria Used in Developing the Tabulation Guidelines

The interagency tabulation working group generated criteria that could be used both to evaluate the technical merits of different bridging procedures (see Chapter 5 and Appendix C) and to display data under the 1997 standards. The relative importance of each criterion will depend on the purpose for which the data are intended to be used. For example, in the case of bridging to the 1977 standards, the most important criterion is "measuring change over time," while "congruence with respect to respondent's choice" will be more critical for presenting data under the 1997 standards.

The criteria set forth below are designed only to assess the technical adequacy of the various statistical procedures. The first two criteria listed below are central to consideration of bridging methods. The next six criteria apply both to bridging and long-term tabulation decisions. The last criterion is of primary importance for future tabulations of data collected using the 1997 standards.

Bridging:

Measure change over time. This is the most important criterion for bridging because the major purpose of any historical bridge will be to measure true change over time as distinct from methodologically induced change. The ideal bridging method, under this criterion, would be one that matches how the respondent would have responded under the 1977 standards had that been possible. In this ideal situation, differences between the new distribution and the old distribution would reflect true change in the distribution itself.

Minimize disruptions to the single race distribution. This criterion applies only to methods for bridging. Its purpose is to consider how different the resulting bridge distribution is from the single-race distribution for detailed race under the 1997 standards. To the extent that a bridging method can meet the other criteria and still not differ substantially from the single-race proportion in the ongoing distribution, it will facilitate looking both forward and backward in time.

Bridging and future tabulations:

Range of applicability. Because the purpose of the guidelines is to foster consistency across agencies in tabulating racial and ethnic data, tabulation procedures that can be used in a wide range of programs and varied contexts are usually preferable to those that have more limited applicability.

Meet confidentiality and reliability standards. It is essential that the tabulations maintain the confidentiality standards of the statistical organization while producing reliable estimates.

Statistically defensible. Because tabulations may be published by statistical agencies and/or provided in public use data, the recommended tabulation procedures should follow recognized statistical practices.

Ease of use. Because the tabulation procedures are likely to be used in a wide variety of situations by many different people, it is important that they can be implemented with a minimum of operational difficulty. Thus, the tabulation procedures must be capable of being easily replicated by others.

Skill required. Similarly, it is important that the tabulation procedures can be implemented by individuals with relatively little statistical knowledge.

Understandability and communicability. Again, because the tabulation procedures will likely be used, as well as presented, in a wide variety of situations by many different people, it is important that they be easily explainable to the public.

Future tabulations:

Congruence with respondent's choice. Because of changes in the categories and the respondent instructions accompanying the question on race (allowing one or more categories to be selected), the underlying logic of the tabulation procedures must reflect to the greatest extent possible the full detail of race reporting.

CHAPTER 2

COLLECTING DATA ON RACE AND ETHNICITY USING THE 1997 STANDARDS

This chapter provides guidelines for use by agencies in developing data collection questions, formats, and associated procedures to implement the 1997 standards.

A. Developing Procedures for Data Collection

An interagency committee representing the Departments of Health and Human Services, Commerce, Education, Labor, and Veterans Affairs, and the General Accounting Office conducted two phases of cognitive research to develop and test procedures to collect and aggregate data on race and ethnicity using the 1997 standards. This chapter briefly describes the research conducted by the committee and offers initial guidelines for agencies developing data collection procedures. These guidelines will be continually reviewed and modified as implementation of the 1997 standards occurs, feedback from agencies is received, and new research findings become available. The guidelines in this chapter address the wording and format of questions that ask for self-reported data on race and Hispanic or Latino origin as well as the design of forms that collect aggregate data on race and Hispanic or Latino origin. Instructions and training procedures for field interviewers and administrative personnel who will be using these questions and forms are also discussed.

1. Developing and Testing Self-Reported Race and Ethnicity Questions

A goal of this research was to provide guidance on the wording and format of questions for self-reported race and Hispanic or Latino origin, depending on the data collection mode. The interagency committee conducted research on survey questions administered by telephone or in

face-to-face personal interviews. In addition, the Census Bureau conducted extensive research on the design of questions in preparation for Census 2000.

Both short and long versions of questions were tested. For short versions of the race question, the five minimum response categories were used–they are American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. Long versions of the race question provided for reporting of subgroups such as Chinese, Japanese, Samoan, and so forth.   For Hispanic or Latino origin questions, the minimum level of detail used was a *Yes* or *No* response indicating Hispanic or Latino origin background. Long versions of the question provided for reporting of subgroups such as Puerto Rican, Cuban, and Mexican.

Below is a brief description of the methods used in the two phases of research conducted by the interagency committee followed by the results and their implications. This section concludes with general guidelines to use in collecting self-reported data on race and ethnicity and offers specific examples of question wording and format.

Research Methods. In Phase I, 44 cognitive laboratory interviews were completed, 33 face-to-face and 11 by telephone. Interviews were conducted in the Washington, D.C. metropolitan area. Subjects were grouped for analysis purposes according to their reports of the race of their mother and father. Among the 44 subjects, 15 reported both parents as Black or African American, 10 reported both parents as White, 2 reported both parents as Asian, 2 reported both parents as Native Hawaiian or Other Pacific Islander, 2 reported both parents as American Indian or Alaska Native, 6 reported their mother's race as different from their father's race, and 7 reported some other response (e.g., Hispanic or country of origin). Of the 6 subjects who reported multiple race backgrounds, 3 reported American Indian or Alaska Native in combination with either Black or African American or White, 2 reported Native Hawaiian or Other Pacific Islander in combination with either Asian or White, and 1 reported Asian and White. Ten of the 44 subjects were of Hispanic or Latino origin.

In Phase II, a total of 82 cognitive interviews were conducted in four locations: New York, NY; Tulsa, OK; Sacramento, CA; and Honolulu, HI.   Half of the interviews tested items designed for face-to-face surveys and the remaining half tested items designed for telephone administration. As in Phase I, subjects in Phase II were grouped for analysis purposes according to their reports of the race(s) of their mother and father. One of the purposes of the Phase II research was to test race and ethnicity questions specifically with subjects who were Hispanic, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and of multiple race backgrounds. Among the 82 subjects, 17 subjects reported their parents as Hispanic; these subjects were not further categorized by race for analysis purposes. Sixteen reported both parents as Asian, 16 reported both parents as Native Hawaiian or Other Pacific Islander, 14 reported both parents as American Indian or Alaska Native, and19 reported their mother's race as different from their father's race. During the cognitive interviews, subjects were probed extensively about their racial and ethnic backgrounds. Based on this information, among the 19 subjects reporting more than one race, 5 were American Indian and White, 3 were Black and

White, 3 were American Indian or Alaska Native, Black, and White, 2 were Asian and White, 2 were American Indian or Alaska Native and Black, 2 were Native Hawaiian or Other Pacific Islander and Asian, 1 was Native Hawaiian or Other Pacific Islander and White, and 1 was Native Hawaiian or Other Pacific Islander, Asian, and White.

All research subjects were asked general demographic questions (e.g., age, education, and marital status) as well as the test versions of questions on Hispanic or Latino origin and race. Respondents were also asked to provide proxy data for all members of their household. Then, debriefings were conducted to learn more about the subjects' understanding of the questions and terms used.

Findings. Generally, subjects were able to answer without difficulty the race and Hispanic or Latino origin questions. In the cognitive interviews, understanding of the intent of a race or Hispanic origin question was shared. However, individual differences were found in the interpretation and meaning of terms used and there was obvious confusion among some subjects regarding the separation of Hispanic or Latino origin from race. In debriefings, some subjects who were of Hispanic or Latino origin said they usually report *Hispanic* (or some variation indicating Hispanic status or country of origin) when answering surveys or government forms that ask race.

As expected, subjects who were interviewed face-to-face seemed to use and rely on the flashcards to select a response. Subjects interviewed by telephone had more difficulty answering the race question and the long version of the Hispanic question since they had to listen to a relatively long list of response options. There was some indication that hearing a list with alternative terms representing one category (i.e., Black or African American is one category, not two) may result in confusion. A few subjects thought the interviewer asked them to choose between Black <u>or</u> African American and commented that they did not like having to make a choice. This problem can be addressed through interviewer training that teaches the interviewer to pause longer after saying each response category; that is, if the interviewer is reading a list of *"...White, Black or African American, Asian, ..."* she/he should pause between the words *White* and *Black*, not pause between *Black* or *African American*, and pause again between *African American* and *Asian*. Last, there was some evidence that hearing the instruction to *"Select one or more..."* was misunderstood on the telephone to mean that the subject <u>had</u> to select more than one race. Interviewers will need to be trained to perceive and correct for this.

Implications. As has been noted elsewhere in the literature, respondents often do not make clear distinctions among the terms and concepts used in defining race, ethnicity, nationality, and ancestry. In the cognitive interviews, understanding of the intent of a race or Hispanic origin question was shared but individual differences in the interpretation and meaning of terms used were found, as was confusion regarding the distinction of Hispanic or Latino origin from race. The following statements from the cognitive interviews illustrate these findings.

C   *It means ethnic background. Not the country. I think people tend to cross quickly between
     using the terms race and country. When I say, "Yes, I am Hawaiian," I mean that in my
     bloodstream I have Hawaii. My blood inheritance.*

C   *Race I guess means the color somebody is. Or, their cultural heritage.*

C   *The word race means the biological heritage from which you descend.*

C   *Race means the culture that someone is from.*

C   *The way I think of race, I think of it as a negative, probably because of what we've read
     about in the 60's--race riots, etc. It always seems to have a negative connotation. I prefer to
     use ethnicity.*

C   *I answer differently sometimes, depending on what's beneficial to my family or me.*

C   *Sometimes you see Hispanic as a choice for race. If Hispanic had been offered as a race
     then I would have chosen that.*

C   *The race question is difficult because it doesn't have enough categories, it's too restrictive.
     With only five categories, there are two that are too specific--American Indian and Native
     Hawaiian--and there's a list of countries for the Asians. It doesn't specify anything about
     Central or South American descent. Everybody comes from different backgrounds; even
     White Americans can probably check off Irish, etc.*

General Guidelines. Based on work accomplished by the interagency committee as well as the
testing of questions in a variety of modes and with subjects of different racial and ethnic
backgrounds, the following guidelines and examples for the design of questions on race and
ethnicity are offered:

C   **Communicate clearly an instruction that allows, but does not require, multiple
     responses to the race question.**

     The 1997 standards are clear that the format and wording used in a question on race must
     communicate to the respondent an instruction that multiple responses are acceptable. Based
     on research findings, the recommended forms for this instruction are *Select one or more,
     Mark one or more,* or *Choose one or more*. There was some limited research indicating that
     the wording *"...one or more..."* was better understood than a *"Mark all that apply"*
     alternative. Other instructions may be needed, especially when integrating a race question
     within an existing data collection instrument. For example, some mail instruments do not
     word questions in a personal way; that is, rather than *What is your age*? an instrument may
     simply have *Age* with a line for an entry. Taking this case further, if a form has an item
     simply worded as *Race* with a line for an entry, then an instruction should be included to

17

communicate that multiple race responses are acceptable (e.g., Race - enter one or more). Regardless of exact wording, the instruction must be evident to the respondent.

C **Consider using an instruction to answer both the question on Hispanic or Latino origin question and the question on race.**

Using an instruction has particular relevance for mail surveys or questionnaires that are self-administered since there is no opportunity for interviewer interaction. An instruction such as the following that was used in Census 2000 may improve potential item non-response, especially among Hispanic respondents: *NOTE: Please answer BOTH Questions 4 and 5 (Hispanic or Latino and Race).*

C **For data collection efforts requiring detailed Hispanic or Latino origin or detailed race information, consider a two-part question or follow-up questions asked by the interviewer or printed on a form.** For example, respondents who first report being of Hispanic or Latino origin would then be asked if they are Mexican, Puerto Rican, Cuban, and so forth.

C **Take mode of administration carefully into account when designing questions and instructions.**

This guideline may seem obvious but it is often the case that surveys are conducted using more than one mode (i.e., the initial interview attempt may be a personal visit but a telephone interview is permissible). Since the questions should be designed with the mode in mind, there may need to be different versions of questions, depending on the mode of administration.

C **Provide definitions to the minimum race categories when possible**.

This guideline is particularly relevant when the short version (only the five minimum categories) of a question on race is used. Individual interpretation of the five categories could lead to response error, especially for respondents unsure of the definitions of Asian, American Indian, Native Hawaiian, and Other Pacific Islander. For self-administered forms, providing the definition of the category should be considered if space and formatting limitations can be overcome. For interviewer-administered questions, the definitions should be readily available to the interviewer (usually in a manual that provides question-by-question specifications or a pop-up screen if the interview is computer-assisted) to assist the respondent if needed.

C **Adhere to the specific terminology for the racial and ethnic categories as stated in the 1997 standards.**

The 1997 standards address the words and terms to use, and also indicate other terms that can be considered. For example, the title of the previous Black category should be revised to Black or African American and additional terms such as Haitian or Negro can be used if desired. In another example, American Indian should be used and Native American should not be substituted for American Indian. Reviewing the terms specified in the revised standards is strongly encouraged before designing questions on race and Hispanic or Latino origin.

Specific Guidelines on Question Wording and Format. The examples below are based on numerous discussions with interagency committee members, recommendations by questionnaire design experts, and testing results from both the interagency committee's research as well as research conducted by the Census Bureau in preparation for Census 2000. It is important to remember that other variations of questions on ethnicity and race may work just as well or better in a particular survey or data collection environment. There is **not** "one right way" to ask an individual to report his/her race and ethnicity. Rather, question wording and format should depend on the mode of administration as well as the context in which the questions are being asked.

For ease of reference, the following list first provides examples for use in a face-to-face or personal visit mode of data collection, followed by telephone and then self-administration (usually thought of as a mail survey, but also could be used for forms and applications filled out by an individual). There are examples of questions that ask Hispanic or Latino origin as well as questions that ask for reports of race. It is important to recognize that as agencies implement the 1997 revised standards, more will be learned about which question formats work best. Thus, OMB does not at this time recommend one example of question wording or format over another. Also, OMB does not recommend a particular order of categories. There are advantages and disadvantages to various approaches such as an alphabetic ordering versus the ordering of the most prevalent group followed by groups less prevalent. The ordering shown reflects the ordering used in the testing of these questions.

Examples of Questions on Hispanic or Latino Origin and Race

Face-to-face administration (assumes flashcards are used in the interview situation)

Example 1   *Are you Spanish, Hispanic, or Latino?*

Example 2   *Are you Hispanic or Latino?*

Example 3   *Are you of Hispanic or Latino origin?*

Example 4   *Are you Spanish, Hispanic, or Latino?*
             If "Yes," ask *Which one of these groups are you? Are you Mexican, Mexican American, Chicano, Puerto Rican, Cuban or of another Spanish, Hispanic, or Latino group?*

Example 5   *Are you Spanish, Hispanic, Latino? For example, Mexican, Puerto Rican, Cuban, or another Hispanic group.*

Example 6   *(Are you/Are any of the persons that I have listed) Mexican, Puerto Rican, Cuban, or of another Hispanic or Latino group?*

Example 7   *Please select one or more of the following categories to best describe your race.*

Example 8   *Please select one or more of the following categories to describe your race.*

Example 9   *Which of these categories best indicates your race? You may choose one or more races.*

Example 10   *Now choose one or more races for each person.   Which race or races does each person consider himself/herself to be?*

Flashcards for face-to-face administration

Flashcard 1   **No**   Not Spanish, Hispanic, Latino
              Yes   Spanish, Hispanic, Latino
                            Includes Mexican, Mexican American, Chicano,
                            Puerto Rican, Cuban, or other Spanish, Hispanic, Latino

Flashcard 2   No, not Spanish/Hispanic/Latino
              Yes, Mexican, Mexican Am., Chicano
              Yes, Puerto Rican
              Yes, Cuban
              Yes, other Spanish/Hispanic/Latino–Specify group

Flashcard 3   White
              Black or African American
              American Indian or Alaska Native
              Asian
              Native Hawaiian or Other Pacific Islander

Flashcard 4   **White**
              Black or African American
              American Indian or Alaska Native
              Asian
                     Asian Indian   Japanese
                     Chinese   Korean
                     Filipino   Vietnamese
                     Other Asian

Native Hawaiian or Other Pacific Islander
>> Native Hawaiian
>> Guamanian or Chamorro
>> Samoan
>> Other Pacific Islander

Flashcard 5   You may choose one or more of the following:
- White
- Black or African American
- American Indian or Alaska Native
- Asian Indian
- Chinese
- Filipino
- Japanese
- Korean
- Vietnamese
- Other Asian
- Native Hawaiian
- Guamanian or Chamorro
- Samoan
- Other Pacific Islander

Telephone administration

Example 11   *Are you Spanish, Hispanic, or Latino?*

Example 12   *Are you Hispanic or Latino?*

Example 13   *Are you of Hispanic or Latino origin?*

Example 14   *Are you Spanish, Hispanic, or Latino?*
If "Yes," ask *Which one of the following are you? Are you Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or of another Spanish, Hispanic, or Latino group?*

Example 15   *(Are you/Is...) Spanish, Hispanic or Latino?* READ IF NECESSARY: *For example, Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or another Spanish, Hispanic, or Latino group?*

If "Yes," ask *Which one of the following Spanish, Hispanic, or Latino groups (do you/does...) identify with? Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or another Spanish, Hispanic, or Latino group?*

If "Other Spanish, Hispanic, or Latino group," ask *What is the name of the other Hispanic group?*

Example 16 *I'm going to read a list of racial categories. Please select one or more to describe your race. Are you White, Black or African American, American Indian or Alaska Native, Asian, or Native Hawaiian or Other Pacific Islander?*

Example 17 *I'm going to read a list of race categories. Please choose one or more categories that best indicates (your/...'s) race. (Are you/Is...) White? Black or African American? American Indian or Alaska Native? Native Hawaiian? or Other Pacific Islander?*

If American Indian, ask *What is the name of your enrolled or principal tribe?*

If Asian, ask *To what Asian group do you belong?* READ CATEGORIES. *Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, or Other Asian?*

If Other Asian, ask *To what other Asian group do you belong?*

If Pacific Islander, ask *To what Pacific Islander group do you belong?* READ CATEGORIES. *Guamanian or Chamorro, Samoan, or Other Pacific Islander?*

If Other Pacific Islander, ask *To what Other Pacific Islander group do you belong?*

Self-administration

Example 18 **Are you Spanish/Hispanic/Latino?**
- ❾ Yes
- ❾ No

Example 19 **Are you Hispanic or Latino?**
- ❾ Yes
- ❾ No

Example 20 **Are you of Hispanic or Latino origin?**
- ❾ Yes
- ❾ No

Example 21 **Are you Spanish/Hispanic/Latino? Mark : the "No" box** if not Spanish/Hispanic/Latino.
- ❾ **No**, not Spanish/Hispanic/Latino
- ❾ Yes, Mexican, Mexican American, Chicano
- ❾ Yes, Puerto Rican
- ❾ Yes, Cuban
- ❾ Yes, other Spanish/Hispanic/Latino - *Print group*
  _____

xample 22   **Are you Spanish/Hispanic/Latino? Mark : the "No" box** if not
Spanish/Hispanic/Latino.

9   **No,** not Spanish/Hispanic/Latino   9   Yes, Puerto Rican
9   Yes, Mexican, Mexican Am, Chicano   9   Yes, Cuban
9   Yes, other Spanish/Hispanic/Latino - *Print group*
_____

Example 23   **Are you Hispanic or Latino?**
_____No, not Hispanic or Latino.
_____Yes, Hispanic or Latino: a person of Cuban, Mexican, Chicano, Puerto Rican, South
or Central American, or other Spanish culture or origin, regardless of race.

Example 24   **What is your race? Mark : one or more races** to indicate what you consider yourself
to be.
9   White
9   Black or African American
9   American Indian or Alaska Native
9   Asian
9   Native Hawaiian or Other Pacific Islander

Example 25   **What is your race? Mark : one or more races** to indicate what you consider yourself
to be.
9   White
9   Black or African American
9   American Indian or Alaska Native — *Print name of enrolled or principal tribe*
_____
9   Asian Indian   9   Native Hawaiian
9   Chinese   9   Guamanian or Chamorro
9   Filipino   9   Samoan
9   Japanese   9   Other Pacific Islander —
9   Korean   *Print race*
9   Vietnamese   _____
9   Other Asian - *Print race*
_____

Example 26   **What is your race? Mark : one or more races** to indicate what you consider yourself
to be.
9   White
9   Black or African American
9   American Indian or Alaska Native - Print name of enrolled or principal tribe
_____
9   Asian Indian   9   Japanese   9   Native Hawaiian
9   Chinese   9   Korean   9   Guamanian or Chamorro
9   Filipino   9   Vietnamese   9   Samoan
9   Other Asian — Print race   9   Other Pacific Islander —Print race
_____   _____

23

Example 27   **What is your race? You may select one or more races.**
    ____   White: a person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

    ____   Black or African American: a person having origins in any of the black racial groups of Africa.

    ____   American Indian or Alaska Native: a person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.

    ____   Asian: a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.

    ____   Native Hawaiian or Other Pacific Islander: a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.

2. Developing and Testing Aggregate Reporting Forms

Implementing the 1997 standards will cause fundamental changes to the ways in which data on race and Hispanic or Latino origin have previously been aggregated and reported. As a result, a second goal of the interagency committee's research is to provide guidance on the design of reporting forms that will be used by administrative personnel to aggregate data on race and Hispanic or Latino origin for a given population (e.g., reporting race and ethnicity for a school population).

Initial research efforts. Three different types of forms were tested with eighteen subjects who were familiar with reporting aggregate data for a given population, but not necessarily familiar with the 1997 standards. Of the 18 respondents interviewed, 6 worked for the Federal Government, 8 worked in private industry, 3 worked in local correctional facilities, and 1 worked in a school.

None of the forms tested were completed accurately without interviewer intervention. Regardless of the form tested or whether the testing was conducted in a laboratory or on-site, the most common problem was the requirement to count and report race for individuals who are of Hispanic or Latino origin. As an illustration, one subject stated *"It's (the form) basically asking how Hispanics were separated into groups of races. I think the part that confuses me is that our Hispanics do not view themselves as another race. And so that is kind of what threw me off… it's asking for Hispanics who had marked 'White,' but they don't. They would have checked Hispanic."* Discussions with subjects revealed that all but one worked for agencies that have used a single question -- a combined race and ethnicity format -- to collect data.

Rather than continuing the testing of different draft forms, work shifted in FY 2000 to establishing guidance for Federal enforcement agencies that collect, use, and/or report aggregate data on race. This work culminated in the March 9, 2000, issuance of OMB Bulletin No. 00-02, Guidance on Aggregation and Allocation of Data on Race for Use in Civil Rights Monitoring

and Enforcement (see Chapter 4 for further discussion on the implementation of this bulletin). The aggregation method described in the bulletin keeps intact the five single race categories and includes the four double race combinations most frequently reported in recent studies. The method also provides for the collection of information on any multiple race combinations that comprise more than one percent of the population of interest. A balance category is provided to aggregate and report those individual responses that are not included in (1) one of the five single race categories or four double race combinations or (2) other combinations that represent more than one percent of the population in a jurisdiction. Appendix B contains the bulletin and an example agencies could use to design aggregate reporting forms.

Guidelines. Even though there were many problems found in developing and testing aggregate forms, the following initial guidelines can be offered:

- **If possible (notwithstanding confidentiality and disclosure issues), allow for the reporting of every combination of multiple race responses.**

- **If every combination cannot be reported because of burden and/or confidentiality concerns, include at a minimum the following 10 categories described in Bulletin 00-02**.

  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
  - American Indian or Alaska Native *and* White
  - Asian *and* White
  - Black or African American *and* White
  - American Indian or Alaska Native *and* Black or African American
  - Balance of individuals reporting more than one race

- **If the categories described in Bulletin 00-02 are used, also include to the greatest extent possible any aggregate counts of multiple race combinations that are greater than one percent of the population of interest or study.**

- **Provide definitions that assist in understanding the concepts of single race responses and multiple race responses as well as the distinction between ethnicity and race.**

- **Explain how the missing data should be reported.**

- **Design the form in a professional manner and include clear instructions.**

- **When feasible, consider providing information to respondents about how multiple race responses will be aggregated and reported to a Federal enforcement agency. For example, employers may want to include on employment applications the following information:**

  Below are two questions--the first is about your ethnicity and the second your race. You are to answer both questions. In answering the second question, you may select one or more races. The summarized information is reported to the Federal Government for civil rights enforcement purposes. The summarized information will be reported in the following categories only:

  1. White
  2. Black or African American
  3. American Indian or Alaska Native
  4. Asian
  5. Native Hawaiian or Other Pacific Islander
  6. Black or African American *and* White
  7. Asian *and* White
  8. American Indian or Alaska Native *and* White
  9. American Indian or Alaska Native *and* Black or African American
  10. Balance of all other individuals selecting more than one race.

  If you select races that are not in categories 1 through 9 above, you will be counted in 10, which is the balance of all other individuals selecting more than one race. For example, if you select Asian *and* Black *and* White, your race will be reported in the balance category.

3. Developing Field Instructions and Training Procedures

Work to develop interviewer instructions and interviewer training procedures has only recently begun. Long-term plans include developing and testing different training modules and interviewer instructions, depending on the mode of administration and the type of data collection. This work will, in all likelihood, address in a more systematic way some longstanding issues in the fielding of questions on race and ethnicity and ways that interviewers can be trained to improve data quality. Specific procedures on how to ask the questions and, in some cases, how to instruct the respondent to use the flashcard, will be developed along with suggested interviewer probes, definitions, and statements that can be used to address respondent problems.

During the cognitive testing of the race and ethnicity questions, interviewer training and field procedures were also tested. Specifically, interviewers were trained to administer the questions in a standardized manner as would be done in actual survey interviews. (Cognitive probing was conducted after the questions were administered.) As is common in actual surveys, the interviewers were also supplied with a "Question and Answer" sheet to assist in responding to questions and confusion on the part of respondents.

The interviewers experienced the greatest difficulty in assisting Hispanic or Latino respondents who were having trouble answering the question on race. If a respondent answered "I'm Hispanic or Latino" (or some other term for Hispanic), the interviewers were trained to point the respondent back to the race categories by responding with something like "In addition to being Hispanic, can you describe yourself as [repeat race categories]?" Also, if a respondent insisted that Hispanic or Latino was a race or asked why it wasn't on the list of races, the interviewer was trained to say that "Hispanic or Latino is generally considered an ethnicity rather than a race–Hispanic or Latino persons can be of any race." This study found these kinds of interviewer explanations to be mostly ineffective and in fact, some respondents found them to be offensive. The research team concluded that interviewers should rely on standard probing techniques to encourage respondents to place their answer into an explicit response category, and to repeat the categories if necessary. Clearly, more research is needed in this area as the 1997 standards become more widely implemented.

B. Processing Census 2000 Data Using the 1997 Standards

This section provides an overview of the procedures the Census Bureau is following in editing responses to the Census 2000 race and Hispanic origin questions and imputing responses to these questions for people who did not provide them. A comparison of these procedures with those implemented in the 1990 census is shown at the end of this section.

The process can be divided into five parts: pre-editing procedures, within-household imputation, "hot deck" imputation, substitution, and group quarters editing. The basic philosophy in performing editing and imputation operations is that the Census Bureau has enough subject-matter expertise and access to related information provided by the respondent, by others in the respondent's household, or by other people similar to the respondent to make reasonable imputations of missing responses. In cases where answers are not responsive to the question, they are removed and new responses are imputed. As part of its effort to inform the public about data quality, the Census Bureau publishes the degree to which it imputes responses for each question. In addition, the Census Bureau indicates in its microdata files which responses for a given record have been imputed.

A certain amount of editing of responses occurs during coding operations, which precede the implementation of formal editing and imputation procedures. For example, during the coding of write-in responses to the question on race, coders must determine how national origin or ethnic group responses, such as Jamaican and German, should be coded into racial categories. The Census Bureau developed an approach for assigning race codes to responses that represent national origins or ethnic groups (excluding American Indian and Alaska Native tribes, Asian

subgroups, and Pacific Islander groups) and the results of this approach are reflected in the race code list and in the coding procedures.[1]

1. Pre-editing Procedures

Purpose: To convert input codes into standard three-digit output codes, detect and correct out-of-range values, ensure that no more than eight[2] race codes appear on the edited file, and resolve into one code multiple responses given to the question on Hispanic origin. (Please note that all original responses are preserved on the unedited files.)

Tasks:

Race
- The pre-editing procedures include the following operations to assign three-digit codes for responses to this question:

    - Convert check box marks into corresponding three-digit codes;

    - Ensure that write-in codes obtained from the coding operations are valid;[3]

    - Eliminate duplicate codes; and

    - Remove general codes when more specific codes are provided (for example, if the check box code for American Indian and Alaska Native and a code for a tribe are present, the check box code is eliminated).

---

[1] Data on single ancestry by race from the 1990 census were used to help make decisions about how to code these responses into racial categories. Essentially, if 90 percent or more of people who reported a single, specific ancestry reported in a specific race category in 1990 (for example, 97 percent of people indicating Jamaican ancestry reported as Black in the question on race), then that race is used as the Census 2000 response. This 90-percent rule was not applied to write-in responses of American Indian and Alaska Native tribes, Asian groups, or Pacific Islander groups because the question on race was designed explicitly to obtain these types of responses.

[2] Although it is possible for more than eight race codes to be input during data processing (including coding), no more than eight race codes are kept on the edited files that are used for data products. For example, people may provide multiple ethnic responses (such as German, Italian, Jamaican, and Nigerian) in addition to multiple check box responses and write-in responses of American Indian and Alaska Native tribes or Asian and Pacific Islander groups, possibly resulting in eight or more input race codes. Results from the Census 2000 Dress Rehearsal show that of people reporting more than one race, the overwhelming majority report only two races. Thus, storing up to eight race codes in the output files preserves, as much as possible, the original groups reported.

[3]In rare cases, invalid codes may have been applied inadvertently to a particular write-in response. Although extensive efforts are made to identify and correct these situations, the editing procedures act as a backup system for resolving any remaining problems.

The pre-editing procedures then ensure that no more than eight race codes are sent as output to the edited file. The goal in this process is to retain as much information as possible about American Indian and Alaska Native tribes and about detailed Asian and Pacific Islander groups while, at the same time, preserving reporting of all other major racial groups such as White, Black, and Some other race.

Hispanic origin

Pre-editing procedures for the Hispanic origin question are considerably simpler because there is only one write-in space and respondents are not asked to report multiple origins. The philosophy of the procedures, however, is similar to that for the race question. Some specific examples include:

- Convert check box marks into corresponding three-digit codes;

- Ensure that write-in codes are valid;

- Override the general code for the "Other Spanish/Hispanic/Latino" check box with the specific code for any origin that is written in. For example, the code for a write-in response of "Guatemalan" (code 222) replaces the check box code for "Other Spanish/Hispanic/ Latino" (code 280); and

- Reduce multiple check boxes marked for a respondent to one output code.[4]

For research purposes, all responses, including reporting of multiple responses, are retained .

2. Within-Household Imputation

Purpose: When race or Hispanic origin data are missing, to impute responses for people from others within the same household who have reported race or Hispanic origin.

Tasks: This part of the editing procedures is performed jointly for the race and Hispanic origin questions. They involve the following steps:

- Identify people in the household for whom no response was given to either or both the Hispanic origin question and the race question;

---

[4] If more than one response was given, obtaining a single response will be achieved as follows: (1) If all the responses are Hispanic, the respondent will be assigned as "Other Spanish/Hispanic/Latino." (2) If all the responses are not Hispanic, the respondent will be assigned as "Not Spanish/Hispanic/Latino." (3) If the responses are a mix of Hispanic and non-Hispanic responses, the responses will be blanked and a single origin will be imputed either, first, from within the household or, if no one in the household gave a single response, from other neighboring households with members of the same race.

- Search to see if an Hispanic origin response can be obtained from a write-in response to the question on race; and

- Search to see if a race response can be obtained from a write-in response to the question on Hispanic origin

After these preliminary steps, the within-household editing procedures follow one of three paths using a predetermined sequence of household relationship to assign race and origin depending on whether both race and origin are blank, only race is blank, or only origin is blank.[5] If both race and origin are blank, the race and origin values assigned will come from the first person in that predetermined sequence with a value for race and/or origin. If only one value is obtained, the procedures for imputing only race (or only origin) are followed.

If only race is blank, the race value is assigned from the first person in that predetermined sequence with the same reported origin group. If only Hispanic origin is blank, the origin value is assigned from the first person in that sequence with the same reported race group.   If race and/or origin cannot be assigned from anyone within the household, then a response is assigned from a "hot deck." (See the next section.)

3. "Hot Deck" Imputation

Purpose: When race and Hispanic origin data are missing from all household records, an origin or race will be assigned from other Census records in surrounding blocks (or nearby households) with "similar" characteristics.

A hot deck is a data table (or "matrix") in which values of reported responses (donors), stratified by selected characteristics of the respondents, are stored and updated on a flow basis and used as needed to assign values of the variable in question to people with similar characteristics who do not have a response. In the case of race, the assignment from the "donor" can be a single or multiple race. Each cell in this table is a "stack" of sixteen stored values that are constantly updated as each household is processed, with the most recently reported value being the first one available for use. Thus, if race cannot be assigned for an individual from within the household, a race is assigned from the first available value in the hot deck "stack" based on age and origin. This value will come from a "donor" with similar age and origin who will have a high likelihood of living nearby, perhaps even next door. Sixteen values are stored in each cell to guard against having to assign the same stored value over and over again if several people in a row with the same characteristics require a hot deck allocation.

---

[5] For example, if only race is needed to be imputed for the child of the householder, the editing procedures would look in the household for the person with a race reported and with the same origin as the child. Records would be searched in the following priority sequence: householder, another child, and spouse of householder.

Tasks: In the race/Hispanic origin editing procedures, there are seven hot deck allocation matrices. Three of them are concerned with allocating both a race and an origin. All three matrices are stratified by three broad age groups (15-34, 35-54, and 55 and older) and focus solely on assigning race and origin to the householder. Use of these matrices occurs when no one in the household has either a reported race or a reported origin. The race and origin assigned to the householder will also be assigned to all other members of the household. Each matrix has a different universe of "donors." The universe for the first matrix will be donors with Spanish surnames (that is, only householders with Spanish surnames are used to update this matrix). The universe for the second matrix will be donors with non-Spanish surnames. The universe for the third matrix will be donors whose names are not clearly either Spanish or not Spanish, or who have not provided a surname on the census form.

The remaining four hot deck matrices assign either race or origin. They are all stratified by four broad age groups (0-14, 15-34, 35-54, and 55 and older). The one matrix that assigns race alone is further stratified by seven origin groups (Not Hispanic and six Hispanic groups: Mexican, Puerto Rican, Cuban, Central American/Dominican, Latin/South American, and Other Hispanic). The three remaining matrices that assign origin alone are stratified by six race groups (White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and Some other race). The universes for the three origin hot deck matrices are, respectively: donors with a Spanish surname, donors with a non-Spanish surname, and donors whose names are not clearly either Spanish or not Spanish or who have not provided a surname on the census form.

4. Substitution

Purpose: To assign characteristics for members of occupied housing units for which there is nothing but a count of people and there are no characteristics reported for anyone in the housing unit.

Tasks: For housing units that are not vacant but for which there are no data, the Census Bureau uses a hot deck technique called "substitution" to assign characteristics (including race and origin) to the people in the housing unit. The assignment of characteristics is achieved through the use of a substitution hot deck matrix which contains "cells" of characteristics for reported households and is stratified according to the type of enumeration method used (mail out/mail back or enumerator) and the number of people in the household. These cells are updated using the characteristics of the most recently reported household of the specified enumeration method and size. As with the other hot deck matrices, each cell stores information for eight households and these cells are constantly being refreshed as new households enter the editing program and are eligible to update the matrix.

5. Group Quarters Editing

Purpose: To assign characteristics to people in group quarters.

Tasks: A separate editing procedure is used for the group quarters population. This editing procedure is necessarily different from the household editing procedure because, in general, people in group quarters are not related to each other and assigning values for members of households depends in some way on household relationships.

For people who do not report an origin, the group quarters editing procedure first searches to see if an Hispanic origin can be obtained from a write-in response to the race question. If not, a group quarters hot deck matrix is used. Each hot deck matrix for assigning origin is stratified by type of group quarters (13 types) and by race, using the same six race categories as those used in the household hot decks. There are three hot deck matrices for origin: (a) one with donors having a Spanish surname, (b) one with donors having a non-Spanish surname, and (c) one with donors whose names are not clearly either Spanish or not Spanish or who have not provided a surname on the census form.

If race is not reported, it will be assigned from the race group quarters hot deck matrix. The hot deck matrix is stratified by type of group quarters (13 types) and by origin (non-Hispanic and six Hispanic groups). The seven origin groups are the same as those used in the household hot decks.

If both origin and race are missing, they are assigned jointly from one of three race/origin group quarters hot decks: (a) one with donors having a Spanish surname, (b) one with donors having a non-Spanish surname, or (c) one with donors whose names are not clearly either Spanish or not Spanish or who have not provided a surname on the census form. These matrices are similar to the joint race/origin hot deck matrices used in the household editing and they are stratified by type of group quarters (13 types) and age (0-14, 15-34, 35-54, and 55 and older).

Differences Between Census 2000 and 1990 Procedures for Editing
Responses to the Questions on Race and Hispanic Origin

	Census 2000	1990 Census
RACE		
Reporting more than one race	Reporting more than one race allowed - Maximum of eight race codes retained.	Reporting more than one race not allowed - Data capture and data processing did not allow more than one race to be retained.
Use of a race response to the origin question to impute a race	Race responses given to the origin question are used to impute a race.	Not used.
Within-household imputation	Assignment of race based on another person in household (according to a pre-defined priority order of household relationship) *with the same origin.*	Assignment of race based on another person in household according to a pre-defined priority order of household relationship. *No origin match required.*
Hot deck imputation	Assignment based on the race reported for the person *with the same age and origin* whose data were most recently processed.	Assignment based on the race reported for the person whose data were most recently processed. *No age or origin match required.*
HISPANIC ORIGIN		
Reporting of more than one origin	Reporting more than one origin not allowed **-** Resolution to one origin using a set of rules; all responses retained for research purposes.	Reporting more than one origin not allowed**-** Data capture and data processing retained only one origin.
Within-household imputation	Assignment of origin based on another person in household (according to a pre-defined priority order of household relationship) *with the same race.*	Assignment of origin based on another person in household according to a pre-defined priority order of household relationship. *Race match not required.*
Surname-assisted hot decks	Separate hot decks depending on whether the surname is Spanish; not Spanish; not clearly Spanish or not Spanish or not reported.	Separate hot decks not used.
RACE AND HISPANIC ORIGIN		
Within-household imputation	Use of joint assignment of race and origin based on another person in household (according to a pre-defined priority order of household relationship), when neither race nor origin was reported.	Joint race/origin assignment within household not used.
Hot deck imputation	Use of joint race/origin hot decks, differentiated by type of surname, when neither race nor origin was reported.	Joint race/origin hot deck not used.
Stack of stored race and origin values in hot deck	16 race/origin values stored.	8 race/origin values stored.

C. Evaluating Census 2000 Data on Race

For many census data users, both governmental and non-governmental and the private sector, there is a need to understand how the Census 2000 race distributions relate to race distributions from previous censuses and current surveys. Adoption of the 1997 standards resulted in a number of changes in the number and names of racial categories and in the sequencing of questions on ethnicity and race. For the 1990 census there were four racial categories (White, Black or African American, American Indian or Alaskan Native, and Asian or Pacific Islander), whereas for Census 2000 there were five racial categories (White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander). In the 1990 census, the question on race preceded the question on Hispanic origin with two intervening questions. For Census 2000, the question on Hispanic origin was immediately before the question on race with a note to respondents to answer both questions. The most profound change to the standards, however, was that of allowing respondents to report more than one race if they chose to do so.

Data by race from most Federal surveys currently reflect a collection methodology of asking respondents to mark only one racial category. Users of the Census 2000 data on race will need to compare the race distribution from Census 2000 to these other sources. To provide data users with a mechanism to make meaningful comparisons of data collected under the 1977 standards with data that are collected under the 1997 standards, the Census Bureau is undertaking two research projects. The first study was part of the Census 2000 data collection efforts. In Census 2000, an experimental panel of 10,000 housing units replicated the 1990 questions on race and Hispanic origin in the 1990 sequence and used the Census 2000 short form as a control panel. Data from the experimental and control panels will be used to evaluate the combined effects of the changes in question wording, format, content, and design on the quality and content of the data on race.

The second study is currently being planned and is expected to be fielded in summer 2001. In this study, data from Census 2000 will be used to identify households where two or more races were reported for at least one respondent, using both short- and-long form households; differential sampling will be used to ensure that households receiving the long form are over sampled. Stratification and differential allocation of the sample to the strata will be explored to increase the precision of the estimates. Stratification will be based on Census 2000 data on such variables as age, Hispanic origin, race, tenure, and urban and rural geographic concentration. Other pertinent information obtained from an analysis of Census 2000 data on race and from external experts will also be considered in the sample design phase. A split panel design will be used; half of the sample households will be mailed a questionnaire that asks respondents to report a single race, while the other half will be asked to report one or more races. Much like in Census 2000, non-response follow up will be conducted for households that fail to return the questionnaire. To the extent possible, households that have moved since completing the Census 2000 questionnaires will be traced. Results from the respondents will be matched to their Census 2000 responses. Additionally, a follow-up interview will be conducted to ascertain

relevant information to help understand the reporting behavior of respondents who reported two or more races and are now being asked to report a single race or vice versa. Background information on single race respondents will also be obtained.

It is expected that reliable estimates that replicate Census 2000 will be produced for the single race distribution and for the five most frequent combinations of two or more races (White and Black; White and American Indian and Alaska Native; White and Asian; Black and American Indian and Alaska Native; and Black and Asian) at the national level. Synthetic modeling is expected to produce reliable estimates at the state and lower geographic levels.

Like the Census Bureau, other Federal agencies also have plans to evaluate data on race collected using the 1997 standards. Future updates of this guidance will include descriptions of these research projects and, if available, will summarize the research results.

CHAPTER 3

TABULATING DATA ON RACE AND ETHNICITY COLLECTED
USING THE 1997 STANDARDS

This chapter describes plans for tabulating data on race and ethnicity collected under the 1997 standards to meet various Federal needs for these data.

A. Census 2000 Data

The Census 2000 questionnaire provided individuals the opportunity to self-report their racial identity by selecting one or more races. For purposes of Census 2000, in an effort to encourage response to this question and to provide the opportunity to gather data on extremely small groups, OMB approved the use of a sixth category -- Some Other Race -- in addition to the minimum set of five racial categories.

This discussion covers tabulation plans for the six categories of race and the two categories of ethnicity (Hispanic or Latino and Not Hispanic or Latino) and for possible combinations of these racial and ethnic categories. It does not address tabulation plans that are being developed for detailed groups of American Indian, Alaska Native, Asian, Pacific Islander, or Hispanic populations for which information was collected in Census 2000.

1. Protection of Data Confidentiality

To maintain confidentiality as required by law (Title 13, United States Code), the Census Bureau uses a confidentiality edit to ensure that published data do not disclose information about specific individuals or households. The result is that a small amount of uncertainty is introduced into census data for small geographic areas to prevent identification of specific individuals or households.

As with data from the 1990 census, a confidentiality edit will be implemented for data from Census 2000 by selecting a sample of census households from internal census files and interchanging their data with data from other households that have identical numbers of household members, but that are in different locations. The net result of this procedure is that the data user's ability to obtain census data is increased, particularly for small geographic areas and small population groups.

2. Plans for Tabulations by Race and Ethnicity

The plans reflect OMB's preliminary guidelines (See Chapter 1, Section B) on tabulations by race and ethnicity. This discussion of the plans covers the presentation of data on both population totals for racial and ethnic categories and on population characteristics (e.g., age, sex,

educational attainment, labor force status, occupation, and income) for racial and ethnic categories.

Before describing these plans for tabulations by race and ethnicity, it is helpful to describe the maximum number of racial and/or ethnic categories for which data could be provided.

There are 63 potential single and multiple race categories -- 6 categories for those who marked exactly one race and 57 categories for those who marked two or more races. These 57 categories of two or more races include the 15 possible combinations of two races (for example, Asian and White), the 20 possible combinations of three races, the 15 possible combinations of four races, the 6 possible combinations of five races, and the 1 possible combination of all six races.

There are two ethnic categories (Hispanic or Latino, and Not Hispanic or Latino). Thus there are 126 categories (63 x 2) in which the population could be classified by both race and ethnicity.

The 63 mutually exclusive and exhaustive categories of race may be collapsed down to 7 mutually exclusive and exhaustive categories by combining the 57 categories of two or more races. These 7 categories are: White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, Some other race alone, and Two or more races.

Alternative groupings for tabulations by race reflect OMB's preliminary guidelines to show "the total selecting each particular race, whether alone or in combination." In combination literally means "in combination with one or more other races." In this "all-inclusive" approach, tabulations will be shown for each of six categories, which will overlap and will add to more than the total population to the extent that individuals report more than one race. These six categories are: White alone or in combination, Black or African American alone or in combination, American Indian and Alaska Native alone or in combination, Asian alone or in combination, Native Hawaiian and Other Pacific Islander alone or in combination, and Some other race alone or in combination.

As in the case of the 63 racial categories, both tabulations by race of the 7 mutually exclusive and exhaustive categories and tabulations by race alone or in combination could be classified by ethnicity (Hispanic or Latino, and Not Hispanic or Latino).

Because of concerns about the usefulness and reliability of data on population characteristics for small populations, about issues with respect to confidentiality, and about providing data products so voluminous that most data cell values would be zero, the Census Bureau is planning (as it has in previous censuses) to present more geographic detail by race and ethnicity for population totals than for population characteristics (e.g., age, sex, housing tenure, education, and income). (The term population characteristics is used here to include both population and housing characteristics. The characteristics of occupied housing units by race and ethnicity are classified

based on the race and ethnicity of the householder.) Specific plans concerning data on population totals and population characteristics by race and ethnicity are discussed later.

3. Overview of Plans for Data Products

The Census Bureau plans to release a variety of data products from Census 2000 in three different media: Internet, CD-ROM, and paper. These data products will include 100-percent data (based on information collected on all questionnaires, such as age, sex, race, and Hispanic or Latino origin) and sample data (based on information collected only on long-form questionnaires, such as education, occupation, and income). Population totals by race and ethnicity are based on 100-percent data, whereas some population characteristics by race and ethnicity are based on 100-percent and some on sample data. The presentation of data by race and ethnicity planned for eight different aggregated data products and for microdata files is discussed below in detail. The eight aggregated data products are:

> (1) Census 2000 Redistricting Data Summary File (100-percent data)
> (2) Demographic Profile (100-percent and sample data on population characteristics for the total population, but no characteristics data by race or ethnicity)
> (3) Summary File 1 (100-percent data)
> (4) Summary File 2 (100-percent data)
> (5) Quick Tables (100-percent data and sample data)
> (6) Census 2000: Summary Population and Housing Characteristics (100-percent data)
> (7) Summary File 3 (sample data)
> (8) Summary File 4 (sample data)

Plans for aggregated tabulations by race and ethnicity are discussed in the following two sections, first for 100-percent data on population totals and characteristics and then for sample data on population characteristics. Microdata files are discussed in the last section.

4. 100-Percent Data on Population Totals and Characteristics by Race and Ethnicity

Census 2000 Redistricting Data Summary File (100-percent data). This file is referred to hereafter as the PL 94-171 file, by which name it is commonly known. Public Law 94-171 requires that the Census Bureau work closely with the "officers or public bodies having initial responsibility for the legislative apportionment or districting of each state" to determine the specific tabulations needed from the decennial census. Tabulations planned for this file are based on meetings and communications with the Redistricting Task Force of the National Conference of State Legislatures and state-appointed liaisons of the governors and legislatures. During this process, senior officials from OMB, the Voting Rights Section of the Department of Justice, and the Census Bureau consulted with the Task Force and state legislative officials.

As described earlier, a confidentiality edit will be used in Census 2000, as was done in the 1990 census. Thus, as in 1990, there will be no data suppression, and there will be many data cells in the PL 94-171 file with very small values (e.g., zero, one, or two).

The PL 94-171 file will include population totals down to the block level for the 63 racial categories described earlier, along with subtotals for the population of one race, for two or more races, and for each combination of two races, three races, four races, five races, and six races. The PL 94-171 file will include four matrices (one-dimensional statistical tables). The first matrix will show the racial categories just described. The second matrix will show the total Hispanic or Latino population and the racial categories for the population that is Not Hispanic or Latino. The racial and ethnic categories included in these two matrices are shown in Table 1. The third and fourth matrices will repeat the first and second matrices, but limited to the population 18 years and older.

From the data that will be presented on the PL 94-171 File, it will be possible to derive population totals for a race alone or in combination (by addition) and for the Hispanic or Latino population by race (by subtraction).

The PL 94-171 file will be available on the Internet and on CD-ROM.

Demographic Profile (100-percent data). For geographic areas down to the census tract level, the Demographic Profile is designed to provide an overview of 100-percent census data that includes all population and housing topics for which data were collected on a 100-percent basis: sex, age, race, Hispanic or Latino origin, household relationship, and housing occupancy and tenure. Given the limited amount of space to show data on each topic, population totals by race and ethnicity will be limited, and data on population characteristics will be limited to the total population. Population totals will be shown for each of the six races alone, for two or more races, and for each major race alone or in combination (as described earlier), but will not be shown for the 57 specific categories of two or more races. The population total will be shown also for the Hispanic or Latino population.

The Demographic Profile will be available on the Internet and CD-ROM and will be available also on paper for governmental units (including states, counties, incorporated places, American Indian and Alaska Native Areas, and Hawaiian Home Lands).

Summary File 1 (100-percent data). For population totals, data shown down to the block level on the PL 94-171 File (including all 63 racial categories) will be repeated on Summary File 1 (SF-1). Data on population characteristics will be shown on SF-1 for the total population and for nine racial or ethnic groups. In some cases characteristics will be shown down to the block level and in other cases down to the census tract level. The ten groups for which population characteristics will be shown are:

(1) Total
(2) White alone
(3) Black or African American alone
(4) American Indian and Alaska Native alone
(5) Asian alone
(6) Native Hawaiian and Other Pacific Islander alone
(7) Some other race alone
(8) Two or more races
(9) Hispanic or Latino (of any race)
(10) White alone, not Hispanic or Latino

SF-1 will be available on the Internet and on CD-ROM.

Summary File 2 (100-percent data). Summary File 2 (SF-2) will show data on population characteristics, subject to a population threshold, down to the census tract level for a large number of groups. As shown in Table 2, these groups include the 63 racial categories, two or more races, the six races alone or in combination, Hispanic or Latino, and Not Hispanic or Latino for race alone and for race alone or in combination. The population threshold for SF-2 is 100. If the population of a racial or ethnic group is less than 100 in a geographic area (e.g., a county or a census tract), population characteristics for that group for that geographic area will not be included on SF-2. (Even if population characteristics for a specific group in a specific geographic area are not available on SF-2, the population total will be available on SF-1.)

SF-2 will be available on the Internet and on CD-ROM.

Quick Tables (100-percent data). Quick Tables (referred to as Table Shells in the Draft Provisional Guidance of 2/17/99) represent a new data product for Census 2000. Each Quick Table is designed to print on one page (from the Internet) and has a fixed table boxhead and table stub (e.g., showing population by age and sex). The Quick Tables, which correspond generally to tables in detailed printed report series from the 1990 census, are supported by the summary files (just as summary tape files supported detailed printed report series in 1990). Quick tables show less data than are available in summary files, but in a more user-friendly format, including more derived measures such as percent distributions. Quick Tables will be available down to the census tract level and will show extensive 100-percent data by race and ethnicity, subject to a population threshold of 100, including population characteristics for racial and ethnic groups included in SF-1 and SF-2.

Quick tables will be available on the Internet.

Census 2000 Summary Population and Housing Characteristics (100-percent data).
This is a printed report series (one report per state and a national summary report) that is designed to provide an overview of 100-percent data in a format that facilitates comparison across geographic areas (e.g., all counties in a state). As with the Demographic Profile, some

population totals will be presented by race and ethnicity, as shown in Table 3, but data on population characteristics will be limited to the total population with the following two exceptions. Data on population characteristics will be shown for the American Indian and Alaska Native population in American Indian and Alaska Native Areas and for the Native Hawaiian and Other Pacific Islander population in Hawaiian Home Lands. This printed report series, which corresponds closely to the 1990 census printed report series on Summary Population and Housing Characteristics, will show data in state reports for counties, county subdivisions, and places (both incorporated and unincorporated).

Census 2000 Summary Population and Housing Characteristics will be a printed report series.

5. Sample Data on Population Characteristics by Race and Ethnicity

Summary File 3 (sample data). Sample data (based on information collected only on long-form questionnaires, such as education, occupation, and income) on population characteristics down to the block group level will be provided on Summary File 3 (SF-3). These data will be shown for the total population and for the same nine racial and ethnic groups for which data on 100-percent population characteristics will be shown on SF-1. The data on SF-3 will be shown subject to a population threshold (see earlier discussion of SF-2); however, this threshold has not yet been determined.

SF-3 will be available on the Internet and on CD-ROM.

Summary File 4 (sample data). Sample data on population characteristics down to the census tract level will be provided on Summary File 4 (SF-4). These data will be shown for the same list of racial and ethnic groups for which data on 100-percent population characteristics will be shown on SF-2. The data on SF-4 will be shown subject to a population threshold (see earlier discussion of SF-2); however, this threshold has not yet been determined.

SF-4 will be available on the Internet and on CD-ROM.

Quick Tables (sample data). The earlier description of Quick Tables in conjunction with 100-percent data applies also to sample data. Quick Tables will be available down to the census tract level and will show extensive sample data by race and ethnicity, including population characteristics for racial and ethnic groups included in SF-3 and SF-4. The population threshold for sample Quick Tables will depend on the threshold selected for SF-3 and SF-4.

Quick Tables will be available on the Internet.

6. Microdata Files

Tabulations on population characteristics by race and ethnicity described above are limited to what is planned for aggregated data products. In addition, the Census Bureau will produce

public-use microdata sample (PUMS) files, as was done in 1990, which will include data by race and ethnicity. Plans for PUMs files from Census 2000, including the amount of racial and ethnic detail, are currently being developed. In 1990, in addition to the confidentiality edit described earlier, the PUMS files were stripped of names and addresses, the order of records was rearranged on the file, and a minimum population threshold of 100,000 was used.

In addition, and subject to the Census Bureau's strict confidentiality standards, the Census Bureau plans to make available on the Internet through the American FactFinder, an Advanced Query Function that permits data users to create tabulations to their own specifications based on microdata files. These microdata files, which underlie the 100-percent and sample summary files, are the 100-percent edited detail file (HEDF) and the sample edited detail file (SEDF).

If a data user wants data on population characteristics for a racial or ethnic group for which characteristics are not available in summary files or Quick Tables and for a geographic area for which a PUMS file is not available, it will be possible -- again, subject to strict confidentiality standards set by the Census Bureau -- to obtain these data with a custom tabulation from the Census Bureau. Because of the strict confidentiality standards, the quantity of data that can be obtained will depend on several factors, including the geographic area, the size of the population universe, and the extent of the characteristics' detail.

**Table 3.1. Racial and Ethnic Categories Planned for the PL 94-171 File
(Internet and CD-ROM)**

(See text regarding protection of confidentiality of data from Census 2000.)

Total population

One race

White
Black or African American
American Indian and Alaska Native
Asian
Native Hawaiian and Other Pacific Islander
Some other race

Two or more races

Two races

White; Black or African American
White; American Indian and Alaska Native
White; Asian
White; Native Hawaiian and Other Pacific Islander
White; Some other race
Black or African American; American Indian and Alaska Native
Black or African American; Asian
Black or African American; Native Hawaiian and Other Pacific Islander
Black or African American; Some other race
American Indian and Alaska Native; Asian
American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander
American Indian and Alaska Native; Some other race
Asian; Native Hawaiian and Other Pacific Islander
Asian; Some other race
Native Hawaiian and Other Pacific Islander; Some other race

Three races

White; Black or African American; American Indian and Alaska Native
White; Black or African American; Asian
White; Black or African American; Native Hawaiian and Other Pacific Islander
White; Black or African American; Some other race
White; American Indian and Alaska Native; Asian
White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander
White; American Indian and Alaska Native; Some other race
White; Asian; Native Hawaiian and Other Pacific Islander
White; Asian; Some other race
White; Native Hawaiian and Other Pacific Islander
Black or African American; American Indian and Alaska Native; Asian

Black or African American; American Indian and Alaska Native; Native Hawaiian
and Other Pacific Islander
Black or African American; American Indian and Alaska Native; Some other race
Black or African American; Asian; Native Hawaiian and Other Pacific Islander
Black or African American; Asian; Some other race
Black or African American; Native Hawaiian and Other Pacific Islander; Some other race
American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander
American Indian and Alaska Native; Asian; Some other race
American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander; Some
other race
Asian; Native Hawaiian and Other Pacific Islander; Some other race

Four races
White; Black or African American; American Indian and Alaska Native; Asian
White; Black or African American; American Indian and Alaska Native; Native
Hawaiian and Other Pacific Islander
White; Black or African American; American Indian and Alaska Native; Asian; Some
other race
White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander;
White; Black or African American; Asian; Some other race
White; Black or African American; Native Hawaiian and Other Pacific Islander; Some
other race
White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific
Islander
White; American Indian and Alaska Native; Asian; Some other race
White; American Indian and Alaska Native; Native Hawaiian and Other Pacific Islander;
Some other race
White; Asian; Native Hawaiian and Other Pacific Islander; Some other race
Black or African American; American Indian and Alaska Native; Asian; Native
Hawaiian and Other Pacific Islander
Black or African American; American Indian and Alaska Native; Asian; Some other race
Black or African American; American Indian and Alaska Native; Native Hawaiian and
Other Pacific Islander; Some other race
Black or African American; Asian; Native Hawaiian and Other Pacific Islander; Some
other race
American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander;
Some other race

Five races
White; Black or African American; American Indian and Alaska Native; Asian; Native
Hawaiian and Other Pacific Islander
White; Black or African American; American Indian and Alaska Native; Asian; Some
other race
White; Black or African American; American Indian and Alaska Native; Native
Hawaiian and Other Pacific Islander; Some other race

White; Black or African American; Asian; Native Hawaiian and Other Pacific Islander;
     Some other race
White; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific
     Islander; Some other race
Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian
     and Other Pacific Islander; Some other race
   Six races
     White; Black or African American; American Indian and Alaska Native; Asian; Native
          Hawaiian and Other Pacific Islander; Some other race

               Total population
Hispanic or Latino (of any race)
Not Hispanic or Latino
   One race
          White
          Black or African American
          American Indian and Alaska Native
          Asian
          Native Hawaiian and Other Pacific Islander
          Some other race
   Two or more races
      Two races
          White; Black or African American
          White; American Indian and Alaska Native
          White; Asian
          White; Native Hawaiian and Other Pacific Islander
   (continue with racial categories shown in the first part of this table)

45

Table 3.2. Racial and Ethnic Categories Planned for Showing Population Characteristics in Summary File 2 (Internet and CD-ROM)

(See text regarding protection of confidentiality of data from Census 2000. "In combination" means "in combination with one or more other races.")

NOTE: The categories listed on this page for showing population characteristics overlap and are shown here in five groups for clarity. "A" is the set of 63 racial categories that is described in the text. "B" and "D" are subsets of the total population. "C" is a set of categories that adds to more than the total population because of multiple-race reporting. "E" is a set of categories that adds to more than the population that is Not Hispanic or Latino because of multiple-race reporting.

A. 63 racial categories shown in Table 1.

B. Two or more races

C. *Race alone or in combination:*
    White alone or in combination
    Black or African American alone or in combination
    American Indian and Alaska Native alone or in combination
    Asian alone or in combination
    Native Hawaiian and Other Pacific Islander alone or in combination
    Some other race alone or in combination

D. Hispanic or Latino (of any race)

E. *Ethnicity by race alone and by race alone or in combination:*
     Not Hispanic or Latino
       White alone
       White alone or in combination
       Black or African American alone
       Black or African American alone or in combination
       American Indian and Alaska Native alone
       American Indian and Alaska Native alone or in combination
       Asian alone
       Asian alone or in combination
       Native Hawaiian and Other Pacific Islander alone
       Native Hawaiian and Other Pacific Islander alone or in combination
       Some other race alone
       Some other race alone or in combination

Table 3.3. Racial and Ethnic Categories Planned for Showing Population Totals in Summary Population and Housing Characteristics (printed report series)

(See text regarding protection of confidentiality of data from Census 2000. "In combination" means "in combination with one or more other races.")

Total population by race:
One race
     White
     Black or African American
     American Indian and Alaska Native
     Asian
     Native Hawaiian and Other Pacific Islander
     Some other race
Two or more races

Total population by ethnicity and race:
Hispanic or Latino (of any race)
Not Hispanic or Latino
  One race
     White
     Black or African American
     American Indian and Alaska Native
     Asian
     Native Hawaiian and Other Pacific Islander
     Some other race
  Two or more races

Selected combinations of two races:
White; Black or African American
White; American Indian and Alaska Native
White; Asian
Black or African American; American Indian and Alaska Native

Race alone or in combination:
White alone or in combination
Black or African American alone or in combination
American Indian and Alaska Native alone or in combination
Asian alone or in combination
Native Hawaiian and Other Pacific Islander alone or in combination
Some other race alone or in combination

47

B. Survey and Administrative Records Data

This section applies to the presentation of data collected under the 1997 standards through surveys and administrative records.

Although these proposed tabulation guidelines are applicable in the near term, they are designed to provide a framework that can be expanded in the future as it becomes possible to present more data on multiple race responses. The main guideline is that data should be presented in as much detail as possible (thereby satisfying the *congruence with respondent's choice* criterion), subject to agency criteria for statistical reliability and confidentiality (satisfying the *meet confidentiality and reliability standards* criterion), and thus the amount of detail presented will be a function of sample size and sample design. In addition, to the extent possible, Federal agencies should report data using the same set of standardized categories to facilitate comparisons across subject-matter areas and data systems, thus satisfying the criteria relating to *range of applicability, statistical defensibility,* and *understandability and communicability.*

The decision to revise the policy for the collection of data on race reflects the increasing complexity of our Nation's demographics. As a result, the ways that survey and administrative record data on race are tabulated and analyzed also will become more complex. The proposed guidelines in this section reflect this complexity. Every attempt was made to keep the tables as simple as possible in order to satisfy the criteria *ease of use* and *skill required.* Examples of tabulation strategies are provided and illustrated using data collected as part of the National Health Interview Survey (NHIS), conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention. Since 1976, the NHIS has allowed respondents to report more than one race, but has also asked respondents to indicate the single race with which they most closely identified. The data on race from this survey have been retabulated for illustrative purposes to be as comparable as possible to the categories in the 1997 standards. (The tables in this section are based on data combined from three years of NHIS data. The resulting larger sample size improves the reliability of the estimates and enables more categories to be shown; however, even when combining three years of data on race, counts for some categories cannot be shown due to small sample sizes. In addition, in some tables, cell counts are divided by a constant to illustrate the situation faced by surveys with smaller sample sizes than the NHIS and/or where it is not possible or appropriate to combine more than one year of data.)

As noted above, agencies are to provide as much detail as possible while adhering to their own standards for data quality and confidentiality. Under a typical data quality standard, a table cell cannot be published if its relative standard error (or other measure of dispersion) is larger than some value specified by the agency. Such a cell would be suppressed (withheld from publication). Under a confidentiality standard, a cell value must be suppressed if knowledge of the cell value might enable someone to gain knowledge about one of the respondents contributing data to the cell. If a cell is suppressed to preserve confidentiality, other cells must also be suppressed so the cell value cannot be derived by subtraction. This is called "complementary suppression." In either situation, information on subgroups that cannot appear

separately in the table would be included in appropriate subtotals and/or in the total. (The reader may wish to refer to *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology* for more information concerning the definition of sensitive cells and the selection of cells for complementary suppression. The Statistical Policy Working Papers are available on the Internet at *www.fedstats.gov* -- go to Policy.)

Since agencies do not use a common set of standards for evaluating confidentiality and quality issues, arbitrary cell size criteria were adopted in this report to illustrate how the application of agency quality and confidentiality standards might affect the cells that can be shown in tables. No data for cells of less than 150 are shown in any table. The tables that result give a preview of the distributions that are likely to result from the implementation of the 1997 standards. Note that since the only data being displayed in this report are population counts, it is possible to show more data cells using the 150 criterion than would be the case if the table presented attributes (income, education, health outcomes, etc.) of these groups. In addition, counts are provided for all ages and both sexes. If data are presented for one sex or one age group, as is most often the case, cell sizes would be much smaller and not as many cells could be reported. Individual survey systems will make decisions as to what data can be shown based on the characteristics of each system and the confidentiality and reliability guidelines established for that data system.

It will not be possible to tabulate two types of responses into the categories identified in the standards. The first type of response is when no information on race is provided. In this report the heading "Race Not Reported" is used for this type of response. This response type can be further subdivided according to the reason that no information was obtained -- refusal, don't know, and not ascertained. The second type is when the response that was received does not match any of the standard racial categories. Such responses are tabulated in this report using the heading "Other Race."

Given that sample size will determine what categories can be shown in a given table, a third heading, "Not Tabulated Above," will be used to include either single or "more than one race" categories that are specified in the standard, but that are not large enough to be published separately. Unknowns are included in this category as well. These three special headings ("Race Not Reported," "Other Race," and "Not Tabulated Above") are used in the tables in this section for illustrative purposes. Since these categories are generally of more interest for methodological rather than substantive analyses, most statistical publications will probably use simpler ways of accounting for these types of responses. Strategies for tabulating these kinds of responses will follow agency policy and the analytic objectives of the report.

A remaining issue to be addressed by Federal agencies is the manner in which data are edited and imputed. The rules used to edit and impute respondents' data on race and ethnicity will affect the racial distributions derived from Federal surveys and administrative records. As noted elsewhere in this report, rules for editing and imputation of data on race and ethnicity should be an area of further research and collaboration for Federal agencies, to ensure that the data reported are as comparable as possible.

Since the objective of this section is to illustrate different tabulation strategies, categories with frequencies too small to be shown will not be treated the same way in all of the tables. In some tables, the category is not shown at all in the table stub but the cell value is included in the total and might be included under the heading "Not Tabulated Above" (as described above). In other tables, the category is retained in order to clarify the structure of the table but data are replaced by a "W" to illustrate that they have been withheld from publication for data quality/confidentiality considerations. When the data are replaced by "W," a footnote is used to describe the reason the data are not shown. No guidance is suggested for handling categories too small to be shown in the table. Such decisions will follow agency guidelines and the objectives of the analyses.

In all tables in this section, the "More Than One Race" heading includes respondents who selected more than one of the five basic racial categories in the 1997 standards. Following recommendations in the standards, many data collection systems obtain information on a more detailed set of responses. When surveys collect more detailed information on race than the minimum five racial categories, some persons may indicate that they identify with more than one of the more detailed groups. For example, within the Asian group, respondents might indicate that they are of Chinese and Japanese heritage. These respondents would not be included in the "More Than One Race" heading, but would be included in the total for Asians. In tables where specific Asian heritages are reported, and if sample sizes are sufficient, an additional Asian sub-category could be used to indicate the number of individuals who marked more than one of the detailed Asian categories.

Table 3.4 provides a detailed set of categories for tabulating data on race that reflects the major dimensions of the 1997 standards. Table 3.4 displays the five single categories, includes more detail on the Asian subgroups, and displays a number of multiple-response categories. The specific multiple response categories that could be presented in other data collection systems would be a function of the overall sample size and the regional characteristics of the population where the sample is selected. The detailed subcategories that are presented should support recreating the minimum basic set of racial categories (American Indian or Alaska Native (AIAN); Asian; Black or African American, Native Hawaiian or Other Pacific Islander (NHOPI); and White).

To illustrate the construction of a table based on a survey with a smaller sample size, all cell counts have been divided by 20. Table 3.5 shows a category for each of the five single racial groups in the 1997 standards as well as a "More Than One Race" heading; sample sizes do not permit the presentation of more detailed categories, including single race AIAN and NHOPI and specific multiple race categories. Data are not shown separately for Native Hawaiians and Other Pacific Islanders, the single race category with the smallest frequency, nor for any of the subcategories under "More Than One Race." The category "Native Hawaiians and Other Pacific Islanders" is retained in the table stub and a "W" in the cell indicates that the value is withheld due to sample size. Since the Native Hawaiian or Other Pacific Islander category is the only single race category that cannot be shown, both the number and the percent for the Native

Hawaiian and Other Pacific Islander group are readily obtained by subtraction. This would be appropriate only if the cell were being suppressed for data quality concerns (i.e., the value is unreliable). If it were being suppressed for confidentiality concerns, another cell would also have to be suppressed to prevent the cell value from being obtained by subtraction.

As was the case under the 1977 standards, it will often not be possible to tabulate data using all of the categories used to collect the information. If data for one or more of the five minimum racial categories fail the requirements for data quality or confidentiality, standard agency products should include them in an aggregation such as "Not Tabulated Above," or only in the totals rather than combining them with categories that are publishable alone. For example, if the data for Native Hawaiians and Other Pacific Islanders cannot be published separately, these data should not be combined with data in the Asian category (except when such combinations are needed for comparability with data collected under the 1977 standards). Instead, the data on Native Hawaiians and Other Pacific Islanders should be included in the total and either omitted from the detailed tabulations completely, replaced with a symbol and footnoted, or included in a separate heading for all groups not specifically tabulated (i.e., under the "Not Tabulated Above" heading.) This last approach is illustrated in Table 3.6 from Table 3.4 where cell counts have been divided by 60. This table is designed to illustrate what might happen when sample sizes are increasingly smaller and data from even fewer categories can be reliably presented. The American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, "More Than One Race," and "Race Not Reported" categories are not listed separately in Table 3.6, but were included both in the Total and the "Not Tabulated Above" rows.

In order to display as much data as possible, as well as to reflect the complexity of reporting on race, some additional categories may be tabulated and reported along with the basic tabulations. These categories may *not* be mutually exclusive but would combine categories to create useful analytic distinctions. For example, a heading could be created for persons reporting that they are Asian whether as a single race or in combination with any other race(s). Parallel categories could be created for any of the five single racial categories. The resulting counts are called "all inclusive." They form distributions for each individual racial group; that is, the sum of the percent of respondents who mark a particular group alone, the percent who mark that group and at least one other group, and the percent who did not mark that group is 100 percent. The all inclusive distributions may provide information on population groups that might not have sufficient size in the sample to be included in basic tabulations. Table 3.7 provides a suggested tabulation strategy. The "NHOPI in combination with other races" category does not meet the criteria for inclusion (n=150) and, therefore, is not shown. If this cell suppression were intended to preserve confidentiality, another cell would have to suppressed as well. This is illustrated by suppressing the "NHOPI all inclusive" category.

Note that when the tabulation involves counts or percentages, the analyst can subtract the count or percentage for each single race from the all inclusive count or percentage to obtain the count of individuals reporting each race in combination with any other race(s). For example, the "Black or African American all inclusive count" minus the "Black or African American" single

race count will yield a count for those reporting "Black or African American in combination with one or more other races." This would not be possible if the tabulation included summary statistics (mean, median, or percent) for attributes such as income, education, or health outcomes.

Tables 3.4 - 3.7 describe tabulation alternatives for data on race collected using the 1997 standards. These standards also affect the collection and reporting of data on Hispanic or Latino origin. The 1997 standards call for asking a question on Hispanic or Latino origin followed by a question on race but also allows under limited circumstances for a single, combined question where Hispanic or Latino origin is included in a list along with the five standard racial categories. In such a combined question, respondents are also instructed to "mark one or more." In either approach, Hispanic origin may be reported alone or in combination with one or more races. As was the case for the tabulation of data on race, data on Hispanic or Latino ethnicity can also be presented for specific subgroups (e.g., Mexican, Cuban, and Puerto Rican) as shown in Table 3.8. The tabulation headings used will be a function of the overall sample size and the population composition where the sample is selected. Table 3.8 presents data only on Hispanic ethnicity and does not include information on race.

Whether separate questions or a combined format is used to collect data on Hispanic or Latino origin and race, there are applications where a combined tabulation of the data on these two dimensions is preferred. Data collected under the 1997 standards using either format will support the analysis of data on both Hispanics or Latinos and non-Hispanics or non-Latinos by race (Table 3.9). For example, Table 3.9 shows that among Hispanics or Latinos, the sample size permits the presentation of data for Blacks, Whites, those of "other" races, and those selecting more than one race. Tabulations which incorporate the Hispanic or Latino subgroup information can be developed by expanding Table 3.9. Since respondents are free to select one or more categories in the combined format, data collected from a survey or administrative reporting system where a combined format is used can also be tabulated using Tables 3.8 or 3.9.

Table 3.4. Sample Tabulation -- Detailed Presentation of Data on Race

Race	N	%
Total	328317	100.00
AIAN	2616	.79
Asian	9718	3.26
Asian Indian	1287	.42
Chinese	2245	.75
Filipino	1965	.63
Japanese	920	.34
Korean	966	.33
Vietnamese	1102	.38
More than one Asian	243	.07
Black	45259	12.32
NHOPI	264	.10
Other	9734	2.22
White	250054	78.24
More than one race	5435	1.62
AIAN/Black	375	.10
AIAN/White	2618	.81
Asian/White	741	.24
Black/White	849	.23
Other/White	277	.08
Race Not Reported	5237	1.45

AIAN   = American Indian and Alaska Native
NHOPI = Native Hawaiian and Other Pacific Islander (for example, Hawaiian, Guamanian, or Samoan)

Note: Not all categories (e.g., Asian subgroups and all possible multiple race groups) are shown due to small cell sizes. Values for these cells are included in the Total category and appropriate subcategories; therefore, subcategories may not add to totals.

Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations

Table 3.5. Sample Tabulation -- Minimum Presentation of Data on Race

Race	N	%
Total	16416	100.00
AIAN	W	.79
Asian	486	3.26
Black	2263	12.32
NHOPI	W	W
Other	487	2.22
White	12503	78.24
More than one race	272	1.62
Race Not Reported	262	1.45

W = Suppressed for data quality or reliability concerns (n<150).
Note: All suppressed cells, including those indicated by a "W," are included in the Total category
and appropriate subcategories.
AIAN   = American Indian and Alaska Native
NHOPI = Native Hawaiian and Other Pacific Islander (for example, Hawaiian, Guamanian, or Samoan)

Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations; cell counts have been
divided by 20 for illustration.

Table 3.6. Sample Tabulation -- Minimum Presentation of Data on Race for a Small Sample

Race	N	%
Total	5472	100.00
Asian	162	3.26
Black	754	12.32
Other	162	2.22
White	4168	78.24
NTA	226	3.96

AIAN   = American Indian and Alaska Native
NHOPI = Native Hawaiian and Other Pacific Islander (for example, Hawaiian, Guamanian, or Samoan)
NTA   = Not Tabulated Above (Includes Race Not Reported, AIAN, NHOPI, and all responses that indicated More Than One Race)

Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations; cell counts have been divided by 60 for illustration.

Table 3.7. Sample Tabulation -- Detailed Presentation of Data on Race and the All Inclusive Distributions.

Race	N	%
Total	328317	100.00
AIAN	2616	.79
Asian	9718	3.26
Asian Indian	1287	.42
Chinese	2245	.75
Filipino	1965	.63
Japanese	920	.34
Korean	966	.33
Vietnamese	1102	.38
Black	45259	12.32
NHOPI	264	.10
Other	9734	2.22
White	250054	78.24
More than one race	5435	1.62
AIAN/White	2618	.81
Asian/White	741	.24
Black/White	849	.23
Race Not Reported	5237	1.45
AIAN all inclusive	5724	1.74
AIAN and other race(s)	3108	.95
Asian all inclusive	10710	3.57
Asian and other race(s)	992	.31
Black all inclusive	46731	12.72
Black and other race(s)	1472	.40
NHOPI all inclusive	402	.14
NHOPI and other race(s)	W	W
White all inclusive	254688	79.65
White and other race(s)	4634	1.41

W = Suppressed for data quality and reliability concerns (n<150). AIAN = American Indian and Alaska Native; NHOPI = Native Hawaiian and Other Pacific Islander (for example, Hawaiian, Guamanian, or Samoan)
Note: Not all categories (e.g., Asian subgroups and all possible multiple race groups) are shown due to small cell sizes. Values for these cells are included in the Total category and appropriate subcategories; therefore, subcategories may not add to totals. Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations

Table 3.8. Sample Tabulation --Hispanic or Latino Ethnicity With Detail

Ethnicity	N	%
Total	328317	100.00
Hispanic/Latino	41585	9.78
Cuban	2151	.54
Mexican	26042	5.86
Puerto Rican	4809	1.25
Not Hispanic/Latino	283735	89.36
Ethnicity not reported	2997	.85

Note: Not all categories (e.g., Asian subgroups and all possible multiple race groups) are shown due
to small cell sizes. Values for these cells are included in the Total category and appropriate subcategories; therefore,
subcategories may not add to totals.

Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations

Table 3.9. Sample Tabulation -- Detailed Presentation of Data on Race and Hispanic or Latino Ethnicity

Ethnicity/Race	N	%
Total	328317	100.00
Hispanic or Latino	41585	9.78
AIAN	391	.09
Asian	334	.09
Black	950	.24
NHOPI	W	W
Other	8348	1.80
White	28742	6.88
More than one race	985	.26
AIAN/White	300	.08
Black/White	163	.04
Other/White	180	.05
Race Not Reported	1816	.42
Not Hispanic or Latino	283735	89.36
AIAN	2160	.69
Asian	9291	3.14
Asian Indian	1263	.42
Chinese	2208	.74
Filipino	1828	.60
Japanese	903	.33
Korean	944	.32
Vietnamese	1082	.47
More than one Asian	216	.08
Black	45259	11.99
NHOPI	226	.10
Other	1303	.41

White	219923	70.96
More than one race	4377	1.35
AIAN/Black	363	.10
AIAN/White	2270	.72
Asian/White	613	.20
Black/White	677	.19
Race Not Reported	2444	.74
Ethnicity Not Reported	2997	.85
Black	298	.08
White	1389	.41
Race Not Reported	977	.29

W = Suppressed for data quality and reliability concerns (n<150).
AIAN   = American Indian and Alaska Native
NHOPI = Native Hawaiian and Other Pacific Islander (for example, Hawaiian, Guamanian, or Samoan)

Note: Not all categories (e.g., Asian subgroups and all possible multiple race groups) are shown due to small cell sizes. Values for these cells are included in the Total category and appropriate subcategories; therefore, subcategories may not add to totals.

Source: NCHS/CDC National Health Interview Survey 1993-1995, Unpublished Tabulations

CHAPTER 4

USING DATA ON RACE AND ETHNICITY COLLECTED UNDER
THE 1997 STANDARDS

This chapter discusses some important uses of data collected under the 1997 standards, reflecting in large measure work that is ongoing.

A. Civil Rights Enforcement and Monitoring

OMB Bulletin No. 00-02. In response to requests from agencies responsible for monitoring and enforcing civil rights laws, OMB led an interagency group that developed guidance on:

- the collection of aggregate data on race when agencies request information from businesses, schools, and other entities and

- the allocation by agencies of multiple race responses, whether individual or aggregate, for use in civil rights monitoring and enforcement.

This guidance (issued as OMB Bulletin No. 00-02, dated March 9, 2000 -- see Appendix B) ensures that agencies can continue to monitor compliance with laws that offer protections for those who historically have experienced discrimination and that reporting burdens are minimized for those reporting aggregate data on race to Federal agencies.

The 1997 standards require, among other things, that agencies offer individuals the opportunity to select one or more races when reporting information on race in Federal data collections. Federal enforcement agencies often collect data on race from businesses, schools, and other entities in aggregate form. To meet these reporting requirements, these institutions will now be collecting single and multiple race responses from individuals and aggregating them for Federal reports.

To simplify and minimize the reporting burden for institutions, an aggregation method has been developed that uses:

- the five single race categories;

- four double race combinations most frequently reported in recent studies;

- any multiple race combinations based on Census 2000 data that comprise more than one percent of the population of interest in the relevant jurisdictions; and

- a balance category to report those individual responses that are not included in (1) one of the five single race categories or four double race combinations or (2) other combinations that represent more than one percent of the population in a jurisdiction.

This method provides consistency across agencies for the reporting of aggregate data, but does not preclude the collection of more detailed information if the agency chooses to do so.

The following *example* illustrates this aggregation method.

1	American Indian or Alaska Native
2	Asian
3	Black or African American
4	Native Hawaiian or Other Pacific Islander
5	White
6	American Indian or Alaska Native *and* White
7	Asian *and* White
8	Black or African American *and* White
9	American Indian or Alaska Native *and* Black or African American
10	> 1 percent: Fill in if applicable_____
11	> 1 percent: Fill in if applicable_____
12	Balance of individuals reporting more than one race
13	Total

To provide consistency across agencies responsible for enforcing civil rights laws, guidance has been developed for how multiple race responses should be allocated. This allocation guidance does not use methods that require either fractional or double counting of individuals, or arbitrary allocation of responses to one minority group versus another. The following rules will be used to allocate multiple race responses for use in civil rights monitoring and enforcement.

- Responses in the five single race categories are not allocated.

- Responses that combine one minority race and white are allocated to the minority race.

- Responses that include two or more minority races are allocated as follows:

  - If the enforcement action is in response to a complaint, allocate to the race that the complainant alleges the discrimination was based on.
  - If the enforcement action requires assessing disparate impact or discriminatory patterns, analyze the patterns based on alternative allocations to each of the minority groups.

Allocation for enforcement purposes should not be confused with various allocation methods for "bridging" to past data collections as discussed in Chapter 5 of this report. The principal purpose of allocation for bridging purposes is to conduct trend or time series analysis.

1. Redistricting

One of the first official statutory uses of data on race and ethnicity collected under the new standards will be for legislative redistricting following Census 2000. The new data format should not require substantial changes in the way redistricting will be conducted.

a. Background

Before a new redistricting plan can legally be implemented, states and political subdivisions that are covered under Section 5 of the Voting Rights Act (which include all state and local jurisdictions in nine states and parts of seven others) are required to demonstrate to either the United States Attorney General or a Federal district court in the District of Columbia that their new redistricting plan has neither the purpose nor the effect of reducing the voting strength of their minority citizens. In addition, all states and political subdivisions are prohibited by Section 2 of the Voting Rights Act from using redistricting plans that have the purpose or effect of discriminating against minority voters, including diluting their voting strength. The U.S. Department of Justice or private citizens may file lawsuits to enforce Section 2.

In order to comply with these Federal laws, states and their political subdivisions typically use the Bureau of Census redistricting count tabulations issued pursuant to Public Law 94-171 to assess the racial and ethnic compositions and distributions of their residents as they draw their redistricting plans. The data are also central to the Department's analysis of whether or not a new redistricting plan reduces the voting strength of minority citizens in violation of Section 5 of the Act. The data are also used for evaluating and proving claims under Section 2 of the Act. For example, the data are used to determine where minority voters are concentrated, as well as to identify areas of racially polarized voting. After the redistricting process is complete, courts rely on the redistricting count data, together with other evidence, to decide any Section 2 legal challenge that may be filed challenging the redistricting plan.

b. How the 2000 Census Data Will Be Used for Redistricting in 2001

The 1990 census Public Law 94-171 tabulations reported data down to the block level for the total population and the voting age population (ages 18 years and older) for four racial groups (American Indian and Alaska Native, Asian and Pacific Islander, Black, and White) and a residual category ("other" race). Data on these racial groups were also cross-tabulated by Hispanic origin. Categories were mutually exclusive (each person was counted only once), and the categories added to the total population reported for a geographic region.

In Census 2000, there are three major changes to the reporting of data on race and ethnicity: (1) the instruction to "mark one or more" racial categories; (2) the splitting of the "Asian or Pacific Islander" category into two separate categories -- "Asian" and "Native Hawaiian or Other Pacific Islander;" and (3) the combining of the 1990 categories for American Indian, Eskimo, and Aleut into a single "American Indian or Alaska Native" category with a write-in space (as in 1990) for principal or enrolled tribe. Hispanic or Latino origin was ascertained in a separate question, as in the 1990 census. The Census 2000 PL 94-171 Redistricting Data will, for the first time, include tabulations of persons who marked two or more races (the multiple race tabulations). In tabulating the PL 94-171 data, the Bureau of the Census will produce 63 tabulation categories: 6 tabulation categories for the 5 "single race" groups plus "some other race"; and 57 tabulation categories for possible combinations of these 6 groups. Such data will be reported in full detail down to the block level to provide the flexibility needed for the enforcement of civil rights programs, and particularly for analysis of redistricting plans under the Voting Rights Act. In accordance with Title 13, U.S. Code, the Census Bureau will continue to protect the confidentiality of individuals during the tabulation and presentation of the Voting Rights data at every step in the process and during the course of Census 2000.

The inclusion of multiple race tabulations in the PL 94-171 redistricting data file has raised questions as to how such data will be analyzed by the Department of Justice pursuant to its responsibilities under the Voting Rights Act. This issue has been addressed in OMB Bulletin No. 00-02, issued on March 9, 2000 (see Appendix B). In its analysis of the Census redistricting data, the Department of Justice will be guided by the bulletin. After aggregating the data pursuant to Part I of the bulletin, multiple race responses will be allocated pursuant to Part II of the bulletin. Thus, any responses that indicate white and one minority race will be allocated to the minority race. Multiple race responses that include more than one minority race will be reviewed to determine if there are any patterns affecting any of the minority races reported. Current research indicates that the number of multiple race responses in the Census 2000 is expected to be small – 2 percent or less. Therefore, the impact of multiple race responses on the analysis of Census data for purposes of the Department of Justice's responsibilities under the Voting Rights Act is likely to be minimal.

The computer record layout reflecting this new design can be accessed now at: http://www.census.gov/clo/www/plrecordnew.pdf. Copies are also available by contacting the Census Redistricting Data Office at 301-457-4039 or at RDO@Census.gov .

2. Equal Employment Opportunity

This section describes the process for implementing the 1997 standards in data collections that are used for monitoring equal employment opportunities.

One of the Federal Government's most significant uses of data on race and ethnicity is in its efforts to ensure that every individual has an equal opportunity for employment. Title VII of the Civil Rights Act of 1964, as amended, prohibits discrimination in employment based upon race, color, sex, religion, national origin, or retaliation/reprisal. Executive Order No. 11246, as amended, similarly prohibits discrimination in employment by Federal Government contractors. Executive Order No. 11246 also requires contractors covered by its provisions to ensure affirmatively that they do not discriminate against their employees and applicants for employment.

Responsibility for enforcement of Federal equal employment opportunity (EEO) laws and regulations is shared among the Equal Employment Opportunity Commission (EEOC), the U.S. Department of Justice, the Department of Labor's Office of Federal Contract Compliance Programs (OFCCP), the U.S. Office of Personnel Management (OPM), and the U.S. Department of Education. The Federal EEO agencies use data on race and ethnicity to ensure nondiscrimination in employment.

In implementing the 1997 standards, the EEO agencies will consider the burden imposed on those required to collect and report racial and ethnic data to the Federal Government. The 1997 standards are not intended to diminish the availability and quality of information collected and made available for civil rights enforcement purposes.

Following the 1990 census, the Census Bureau, under contract to EEO agencies, created an EEO Special File that has been generally used as a benchmark for employment availability determinations. It is anticipated that a similar file containing data collected under the 1997 standards will be created from the Census 2000 data.

a. Employee and Applicant Reporting Requirements – Methods Used Prior to Implementing the 1997 Standards

Data on each private employer's workforce are collected annually on the Employer Information Reports (EEO-1 Report). Biennial surveys are collected from local referral unions (EEO-3), state and local governments (EEO-4), elementary and secondary public schools (EEO-5), and post-secondary schools using the Integrated Post Secondary Education Data System (IPEDS Fall Staff Survey, formerly EEO-6). Currently, these forms collect information about each employer's workforce by job category, gender, and race/ethnicity using the categories in the 1977 standards.

Under the 1977 standards, employers were only required to collect, maintain, and report single race information on employees, using the following five categories: (1) White, not of Hispanic origin; (2) Black, not of Hispanic origin; (3) Hispanic; (4) Asian or Pacific Islander; and (5) American Indian or Alaskan Native. Instructions on the current EEO forms state that the race/ethnicity of an employer's workforce may be obtained either by "visual surveys of the workforce, or from post-employment records." The current instructions explicitly state that eliciting information from employees via direct inquiry is not encouraged.

The OFCCP is developing an Equal Opportunity Survey (EO Survey) to obtain employment information from selected Federal contractor establishments. The EO Survey will collect information about applicant and employee personnel activity, and employee compensation, by gender, race, and ethnicity, sorted by EEO-1 job categories. The survey requests that each applicant or employee be identified by gender, by Hispanic or non-Hispanic origin, and by one of the five racial categories. Contractors responding to the EO Survey have discretion to use the 1977 racial/ethnic categories if they have not yet converted their record keeping to the 1997 standards. Educational institutions, however, should follow the guidelines set forth in Section 3 on Equal Access to Education for reporting data on their employees.

For reporting covering 2000 through 2002, employers (including contractors) and unions subject to these reporting requirements will continue to submit EEO-1, EEO-3, EEO-4, EEO-5, OFCCP EO Survey, and IPEDS Fall Staff Survey, in their respective current formats. Employers will use the new forms that reflect the 1997 standards starting in 2003 to provide data on the work force for the 2003 calendar year.

b. Required Changes to EEO Forms and Instructions to Implement the 1997 Standards

To be consistent with the 1997 standards, certain changes are planned for EEO reporting forms and instructions. These changes (along with updates to reflect changes in the industry and occupational classifications) will be submitted as usual under the Paperwork Reduction Act for OMB approval, allowing sufficient time so that non-Federal employers may prepare for their use starting in 2003.

Data collections covering Federal employment will also be updated by OPM and EEOC to reflect the 1997 standards. OPM is engaged in ongoing discussions with numerous Federal agencies to develop a standardized approach for collecting and reporting data under the 1997 standards.

The following changes are planned for the reporting forms and instructions:

    -- The "Native Hawaiian or Other Pacific Islander" racial category will be added.

    -- "Asian" will replace the category "Asian or Pacific Islander."

-- The following changes in terminology will be made:
  -- "Alaska Native" will replace the term "Alaskan Native."
  -- "Black or African American" will replace the term "Black."
  -- "Hispanic or Latino" will replace the term "Hispanic."

-- Data on Hispanic or Latino ethnicity will be collected in a separate question for each occupational group.

c. EEO Approach to Data Collection and Record Keeping Requirements Under the 1997 Standards

Federal EEO agencies use a variety of reporting forms to collect data from businesses, schools, and other establishments. These forms are subject to review and approval by OMB under the Paperwork Reduction Act. In redesigning EEO forms to comply with the 1997 standards, the following categories for race will be recommended for EEO data collection and record keeping:

American Indian or Alaska Native
Asian
Black or African American
Native Hawaiian or Other Pacific Islander
White
American Indian or Alaska Native *and* White
Asian *and* White
Black or African American *and* White
American Indian or Alaska Native *and* Black or African American
Balance of individuals reporting more than one race

The balance category will be used to report individual responses that are not included in any of the single race categories or in any of the two-race categories listed above.

When Census 2000 data become available, the EEO agencies will undertake an analysis of the distribution of two-race combinations. This analysis will be undertaken, first, to confirm the four most common multiple race combinations found at the national level. If they are different from the four previously mentioned, the EEO proposed forms will be changed accordingly. Second, this analysis will be used to determine whether there are any additional multiple race combinations that should reported in the EEO proposed forms.

In addition, the following changes in EEO data collections are being made as part of the implementation of the 1997 standards:

- Self-identification will be the preferred method of collecting data on race and ethnicity in the employment setting.

- Individuals will be permitted to report one or more races on applications and other forms pertaining to their employment.

- A two-question format will be used to obtain information on the ethnicity and race of each applicant or employee. The ethnicity question (Hispanic or Latino *or* Not Hispanic or Latino) will be asked first, followed by the race question.

- If an applicant or employee declines to self-report, employers may then determine the information on race and ethnicity required for reporting purposes by visual observation or from post-employment records.

d. Ensuring Common Approaches in EEO Enforcement

As set forth by OMB in Bulletin No. 00-02 (see Appendix B), the following principles apply regarding data on race in **enforcing** employment discrimination laws and regulations:

- Individuals who select only one racial category will be counted in that category.

- Individuals who select one minority race *and* White will be counted as part of that minority race.

- For the purpose of analyzing an enforcement action, responses of individuals who select more than one minority race will be allocated to the race involved in the enforcement action, if that race is among the ones they have selected. For example, where an enforcement action is based on complaints of discrimination on the basis of Black or African American, responses from individuals reporting themselves as Black or African American and American Indian or Alaska Native will be allocated into the Black or African American category for the purpose of analyzing the enforcement action.

- For nonenforcement-related analyses covering internal and external workforce distributions, several implementation issues remain unresolved at this time. For example, questions covering increased employer record keeping burdens under alternative allocation approaches must be examined further. Additionally, where the enforcement involves Federal litigation, questions remain to be resolved regarding judicial interpretation of both analytical methodology and admissible evidence standards that cover private and Federal lawsuits.

e. Resurveying the Workforce

In converting to a new EEO-1 form in 2003 that reflects 1997 standards, employers will have to reconcile the 1997 standards with the pre-1997 standards under which employee race and ethnicity will be reported through 2002. In issuing the 1997 standards, OMB did not indicate

that employers, in implementing the new standards, would be required to resurvey incumbent employees whose race/ethnicity was originally classified under the pre-1997 standards.

Each EEO agency will determine if resurveying should be required as part of its enforcement activity. Employers who report to more than one Federal agency, one of which requires resurveying, will be required to resurvey irrespective of whether the other Federal agencies impose the same requirement. After a careful review of the pros and cons associated with resurveying, EEOC may require resurveying as part of its record keeping and reporting requirement. EEOC seeks to ensure that the race and ethnicity reported by employees hired under the old standards prior to 2003, and by employees hired under the 1997 standard beginning with reports filed in 2003, will be consistent and comparable. EEOC is concerned that, unless resurveying is conducted, the racial identity of employees previously reported as Hispanic would be unknown. Complete racial and ethnic comparisons are needed to ensure data continuity and comparability, and to promote accuracy in enforcing EEO laws and regulations.

3. Equal Access to Education

Another significant use of race and ethnicity data occurs in the context of the Federal Government's efforts to ensure that every individual has an equal opportunity for education and to collect and analyze data on the educational experience in America. The Department of Education's Office for Civil Rights (OCR) and the Department of Justice's Civil Rights Division are responsible for enforcement of the statutes that protect equal opportunity in education regardless of race, color, sex, national origin, age, or disability pursuant to Title VI of the Civil Rights Act of 1964. OCR also monitors whether a recipient of Federal financial assistance has utilized criteria or methods of administration that have the effect of subjecting individuals to discrimination on the basis of race, color, or national origin. The Department of Justice also enforces Title IV and Title VI, as well as the Equal Educational Opportunities Act of 1974, and also may intervene in private suits that allege violations of education related anti-discrimination statutes and the Fourteenth Amendment of the U.S. Constitution. In addition, the National Center for Education Statistics is authorized to collect, analyze, and disseminate statistics on race and other population characteristics when such information would facilitate educational policy and decision making.

a. Data Collection and Record Keeping Requirements Under the 1997 Standards

On March 9, 2000, OMB issued Bulletin No. 00-02 (see Appendix B) providing guidance for Federal enforcement agencies in their use of data on race for civil rights monitoring and enforcement. The guidance indicates that for this purpose Federal agencies should aggregate data into the five single race categories, the four most frequently reported multiple race combinations, and any multiple race combinations that comprise more than one percent of the population of interest.

b. Racial and Ethnic Data in Educational Institutions

Data on the race and ethnicity of students are critical to many sound educational practices and to monitoring and enforcing civil rights in educational institutions. The elimination of discrimination in educational institutions ensures that all individuals have equal access to educational opportunities. Although Federal enforcement agencies do not determine civil rights compliance based on numbers alone, analyzing student populations by race and other important factors can identify potential issues or problems. This information also enables schools to take proactive steps to prevent potential civil rights violations. Furthermore, assessing the performance of districts, schools, and students by examining information on race and ethnicity can serve important educational goals. Such assessments can allow schools to identify where new approaches or programs are needed and whether existing programs are serving the needs of all students.

The Department of Education has received information from individual schools and states as well as anecdotal reports in the media that suggest that asking children to self-report their race may lead to higher multiple race reports when compared to adult self-reports. Some information indicates that this rate could be substantially higher, particularly in certain jurisdictions. For example, one state has informed Federal officials that 16 percent of its school age population would identify itself as more than one race (in contrast to the estimated 2 percent of the overall nationwide population). In this state, there is also evidence which suggests that some of these students are inconsistent in identifying their race and ethnicity; that is, they change their answers to race and ethnicity questions depending on the situation and/or over time. Moreover, there is research evidence that as children with multiple racial heritages age, they tend to identify with one race in somewhat larger numbers than when they were younger.

Federal enforcement agencies must balance the interest in obtaining complete, detailed data to facilitate effective monitoring of the civil rights laws with the need to minimize the cost and burden on schools and colleges of collecting and reporting information. The collection and reporting of all possible combinations of race and ethnic categories would create an excessive burden for many, if not most, institutions. Further, such an excessive burden could dramatically slow the receipt of information on educational institutions and thereby delay the enforcement process. The Federal Government recognizes these potential burdens. Bulletin No. 00-02 states that "we must minimize reporting burden for institutions such as schools and businesses that report aggregate data on race to Federal agencies." The Bulletin also notes, however, that "[a]s the revised standards for collecting and presenting data are implemented, we must ensure that we maintain our ability to monitor compliance with laws that offer protections for those who historically have experienced discrimination."

The Department of Education recognizes that educational institutions need to maintain consistency between changes affecting data on students and changes affecting data on staff and faculty. The Department of Education has elected to wait until after Census 2000 data are available to determine how educational institutions should aggregate racial and ethnic data for

their students and employees under the 1997 standards. Based on these data, the Department of Education will determine which race combinations (in addition to the five single race and four largest multiple race combinations) meet the one percent threshold for the relevant jurisdictions. The Department then will provide guidance to educational institutions on how racial and ethnic data should be aggregated and reported.

c. Collection of Data on Race and Hispanic Ethnicity

In addition to collecting data on race, the Department of Education and educational institutions at various state and local levels also collect information on Hispanic ethnicity. The tabulation of information on Hispanic ethnicity and race in light of the 1997 standards raises some challenging questions. As with the collection and aggregation of detailed data on race, the Department of Education recognizes the importance of obtaining more complete Hispanic origin information to facilitate effective civil rights monitoring and enforcement. However, the burden on educational institutions must be considered and balanced.

There are three potential options for a unified tabulation approach for data on race and ethnicity. Some suggest that Hispanic or Latino data do not need to be tabulated by race. This suggestion is supported by an effort to minimize the burden on data collectors as well as evidence that many who identify themselves as Hispanic or Latino do not select a race (estimates range up to 30 percent). Others have suggested that the collection of data on race for those of Hispanic or Latino origin that mirror the non-Hispanic categories would create a symmetry that facilitates understanding and implementation. Alternatively, the collection of Hispanic or Latino data by some but not all racial categories could balance the need for the racial identification of Hispanics for civil rights enforcement with the burden on data collectors. Once the Department of Education reviews the results of Census 2000, the Department will reach a final decision on what data will be collected on the racial identification of Hispanic/Latino individuals.

In general, educational agencies that collect data need three years from these final decisions to reconfigure their systems for the collection of racial and ethnic data under the 1997 categories. In the interim, educational institutions may collect information in greater detail than is presently required as long as any additional categories can be aggregated back into the minimum set of categories.

4. Enforcement of Title VI of the Civil Rights Act of 1964

All agencies that enforce regulations under Title VI of the Civil Rights Act of 1964, as amended, are obligated to follow the guidance provided in this document and OMB Bulletin No. 00-02, which addresses aggregating and allocating data on race for civil rights enforcement and monitoring. The Title VI Coordination Regulations at 28 C.F.R. § 42.406, provide that all agencies "shall in regard to each assisted program provide for the collection of data and information from applicants for and recipients of Federal assistance sufficient to permit effective enforcement of [T]itle IV." All agency Title VI regulations provide for the collection of such

data. Agencies currently follow the categories set forth in the Title VI Coordination Regulations, which "are in conformity with the OMB Ad Hoc Committee on Race/Ethnic Categories' recommendations." See 28 C.F.R. § 42.302(e). The section states, however, that to the extent that these designations are modified by OMB, the regulation is to be interpreted to be consistent with any such OMB modifications. See 28 C.F.R. § 42.402(e)(5). Accordingly, the categories listed in the regulation are no longer in effect.

The 1997 revised standards, effective October 30, 1997, require all Federal agencies with Title VI compliance and enforcement obligations to provide for the collection of data by offering respondents the option of selecting one or more of the following racial categories:

1. American Indian or Alaska Native
2. Asian
3. Black or African American
4. Native Hawaiian or Other Pacific Islander
5. White

Pursuant to OMB Bulletin No. 00-02, agencies are instructed to tabulate data on race using the five single race categories outlined above plus the following four multiple race combinations most likely to occur:

1. American Indian or Alaska Native *and* White
2. Asian *and* White
3. Black or African American *and* White
4. American Indian or Alaska Native *and* Black or African American

Agencies are also advised to tabulate additional multiple race combinations that are greater than one percent of the population at issue and include these combinations as part of any tabulation.

Thus, all agency tabulations for data on race should reflect a *minimum of ten categories*: the five single race categories, the four multiple race categories, and a balance category. In surveys where there are additional multiple race combinations that are greater than one percent of the population at issue, those categories should be included in any final tabulation.

All agencies with Title VI enforcement responsibilities should also be aware that OMB Bulletin No. 00-02 provides guidance for the allocation of multiple race responses for use in civil rights monitoring and enforcement. Pursuant to that guidance, the following rules apply:

• Responses in the five single race categories will not be allocated.

• Responses that combine one minority race and white are allocated to the minority race.

• Responses that include two or more minority races are allocated as follows:

- If the enforcement action is in response to a complaint, a multiple race response will be allocated to the race that the complainant alleges the discrimination was based on.

- If the enforcement action requires assessing disparate impact or discriminatory patterns, analyze the patterns based on alternative allocations to each of the minority groups.

Under the 1997 standards, "Hispanic or Latino" is an ethnic category, not a racial category. Where agencies collect data on race and ethnicity separately, ethnicity must be collected first. In such cases, provisions shall be made to report the number of respondents in each racial category who are "Hispanic or Latino" and who are "Not Hispanic or Latino."

All agencies must modify all new and revised record keeping or reporting forms that include racial and/or ethnic information to conform to the 1997 standards. All existing record keeping or reporting requirements must be consistent with these standards at the time they are submitted for extension, but not later than January 1, 2003. Agencies should note that OMB is the final arbiter of all modifications to racial and ethnic categories used in Federal data collection instruments.

Executive Order No. 12250 gives the Attorney General authority to ensure the consistent and effective enforcement of Title VI and other nondiscrimination statutes that apply to recipients of Federal financial assistance. That authority has been delegated to the Civil Rights Division in the Department of Justice. The Division, through the Coordination and Review Section, will ensure that all agencies are aware of the 1997 standards and are taking appropriate steps to implement the OMB guidance. The Division's Coordination and Review Section will be available to Federal agencies to assist them in this endeavor.

B. Intercensal Estimates and Vital Records

Following some background discussion, this section presents a description of the Census Bureau's Intercensal Population Estimates Program, its data sources, methodology, and major uses, and then discusses some of the important issues that must be addressed in implementing the 1997 standards in this program.

Background. In 1977, the Office of Management and Budget (OMB) issued Race and Ethnic Standards for Federal Statistics and Administrative Reporting. Because the intercensal population estimates are limited in their detail by the availability of administrative data, it was 1993 before the program could implement fully the 1977 standards by providing data for the population in the four racial categories of that standard -- White; Black; Asian or Pacific Islander; and American Indian or Alaskan Native. To comply with the 1977 standards, the Intercensal Population Estimates Program developed estimates by race separately for the population by Hispanic origin (Hispanic, Non-Hispanic).

The 1997 standards present many challenges, with two in particular posing the greatest challenges. One is that respondents to Federal data collections, including Census 2000, surveys, and vital statistics registrations, will be allowed to select one or more races. The other is that the Asian or Pacific Islander category has been split into two categories -- one called "Asian" and the other called "Native Hawaiian or Other Pacific Islander."

The population estimates are data driven. Changes to the program to produce estimates for new racial categories will depend upon the availability of data from a variety of sources. Although changes are possible, discussions with data providers and data users, as well as research and analysis of data collected using the 1997 standards, will be required before the Census Bureau can identify the racial categories that can be used in the Intercensal Population Estimates Program.

Because the intercensal population estimates serve several diverse purposes, exploring the possible outcomes of the estimates process and examining the implications of the 1997 standards are important. The issues raised by the 1997 standards are complicated and diverse. It will take considerable research and experimentation before the Intercensal Population Estimates Program can produce population estimates that fully follow the 1997 standards.

The next sections describe the program and discuss the major issues that must be addressed in changing program outputs.

1. Description of the Intercensal Population Estimates Program

The Intercensal Population Estimates Program, under Title 13, develops and releases annual estimates of the total population and its demographic characteristics. For the Nation, states, and counties, these characteristics include annual estimates by: age (single years of age (age 0 to age 99) and 100+); sex (male/female); race (White; Black; Asian and Pacific Islander; and American Indian, Eskimo, and Aleut); and Hispanic origin (Hispanic/Non-Hispanic).

The Intercensal Population Estimates Program currently provides estimates of the total population of functioning governmental units (cities, incorporated places, and minor civil divisions). The Census Bureau is considering expansion of the program to include smaller and more diverse units of geography (such as school districts), as well as the development of demographic characteristics for functioning governmental units and other smaller geographic units.

2. Uses of Population Estimates

The population estimates are used in the intercensal period for funding allocations, as controls for Census Bureau and other Federal surveys, as denominators for vital statistics and other demographic events, and as planning tools for government and private programs and policy decisions.

Funding Allocations. Federal programs totaling $180 billion use these annual population estimates to make program decisions and to distribute these funds.

Survey Controls. The population estimates are used as control totals for the Current Population Survey (CPS), the Survey of Income and Program Participation (SIPP), the new American Community Survey (ACS), other Federal surveys, and many private sector surveys.

Most Federal surveys use national level population estimates by age, sex, race, and Hispanic origin as controls for weighting survey data. The ACS currently uses county level population estimates by age, sex, race, and Hispanic origin as controls for weighting survey data.

Denominators for Demographic Events. The National Center for Health Statistics (NCHS) currently uses the national, state, and county population estimates by age, sex, race, and Hispanic origin as denominators to create birth and death rates and to calculate life tables by race and sex. The Centers for Disease Control and Prevention (CDC) frequently relies upon the estimates of population at various geographic levels as denominators for various health related and disease incidence rates. The National Cancer Institute (NCI) uses the county population estimates by age, sex, race, and Hispanic origin as denominators for the various cancer incidence rates released to the public.

Planning Tools. The intercensal population estimates are frequently used as planning tools and as barometers to measure an area's growth and change since the last decennial census. In making policy decisions, local planners frequently cite the overall population level and the demographic characteristics products of the Intercensal Population Estimates Program.

3. Methodology for Developing Intercensal Population Estimates

The Intercensal Population Estimates Program develops its population estimates by age, sex, race, and Hispanic origin using the demographically recognized cohort-component technique. In this technique, each component of population change -- births, deaths, international migration, and internal migration -- is estimated separately by age, sex, race, and Hispanic origin. Various administrative records provide information needed to develop these components of population change. The estimates process begins with the most recent decennial census results and combines the estimated components of population change to develop the intercensal population estimates.

The 1990 Census Base Population. Although the enumeration of the resident population in the 1990 census, without adjustment for net undercoverage, was adopted as a standard for the estimates, changes were made in the distribution of the population by age and race. These modifications were made to bring the definition of age and race into conformity with definitions used for data from other sources, such as vital statistics. (See Comparability Issues below for a discussion of the modification of the 1990 decennial census.)

Birth and Death Components. In brief, NCHS provides annual counts and distributions of births and deaths by age, race, sex, and Hispanic origin by county to the Census Bureau in a specially developed individual record file of the birth and death events. These individual records contain the detailed race and Hispanic classifications available from the birth and death certificates collected by NCHS.

International Migration Component. The international net migration components are based on a variety of administrative sources and analytic estimates. The Immigration and Naturalization Service (INS) supplies data on legal immigrants. The Office of Refugee Resettlement (ORR) supplies data on persons admitted to the United States as refugees. Both sources supply data on country of birth. The Census Bureau estimates the distribution by race and Hispanic origin from the country-of-birth tallies, using data from the 1990 Census on the foreign-born population who entered the United States from 1985 to 1990.

The other components of international migration such as emigration and undocumented migration are developed using a combination of basic demographic modeling techniques. By examining data from other administrative records in combination with an analysis of the decennial census, the Census Bureau models the level and demographic characteristics of these other international migration components.

Internal Migration Component. Data on internal migration are developed using a basic administrative records method. This method relies on annual extracts of tax returns provided by the Internal Revenue Service (IRS). In this approach, using the Social Security Number (SSN) on the return, the Census Bureau can match the tax returns for two years and obtain state of residence for the two periods. By comparing the state of residence at the two points in time, annual measures of migration can be developed for states.

Until recently, the Census Bureau had only developed the national population estimates by age, race, sex, and Hispanic origin and the estimates of the total population for states and counties. During the current decade, the Census Bureau started to develop a set of state and county population estimates by age, sex, race, and Hispanic origin.

The state population estimates are developed using the basic cohort component technique outlined above. Since the standard tax return provides no demographic characteristics of the tax filer, the Census Bureau must further modify the basic administrative records method to estimate internal migration by age, sex, race, and Hispanic origin. To obtain demographic characteristics, the Bureau has relied on the annual extract of tax returns provided by the IRS, and a 20 percent sample of information on the Social Security Administration (SSA) Application File (NUMIDENT). This NUMIDENT file includes SSN, month and year of birth, race, sex, and six characters of the last name for each SSN holder in the sample file.

The extract of the NUMIDENT file has been merged with the tax returns file by SSN to derive demographic characteristics of IRS filers. Because the Census Bureau was able to receive only a

20 percent sample of this basic NUMIDENT file, the Bureau appended the demographic characteristics of the primary filer to only the same 20 percent sample of tax returns. Besides demographic characteristics of the primary filers, the model requires demographic characteristics of those persons claimed as exemptions on the tax return. The rules for assigning demographic characteristics to dependents are straightforward and rely on basic familial and demographic relationships.

Because the NUMIDENT File was restricted to a 20 percent sample until this year, the Census Bureau could not use the merged tax file and SSA data to develop county population estimates by age, sex, race, and Hispanic origin. To develop the current sets of county population estimates by age, sex, race, and Hispanic origin, a ratio approach is employed. This approach combines the full set of age, race, sex, and Hispanic origin detail for the county in 1990 with the newly developed state population estimates by age, sex, race, and Hispanic origin and the estimates of the total population of the county. With the delivery of the 100 percent NUMIDENT file to the Census Bureau, work on employing the cohort component technique to develop the county estimates by age, sex, race, and Hispanic origin is anticipated.

4. Data Availability

The intercensal population estimates are "data driven." As noted above, the decennial census, the National Center for Health Statistics, the Immigration and Naturalization Service, and the Social Security Administration are all important sources for developing intercensal population estimates. Using the current methodology, estimates cannot be produced without the availability of these data.

Decennial Census Data. Census 2000 is the first time that decennial population data are available using the 1997 standards for collecting racial data.

Birth and Death Data. The National Vital Statistics System is the basis for the Nation's official statistics on births and deaths (including infant deaths). The data are provided through vital registration systems maintained and operated by the individual states and territories where the original certificates are filed. While the legal authority for vital registration rests with the states and territories, the National Center for Health Statistics (NCHS) is required to produce national vital statistics by collecting data from the vital records of all the states. The NCHS cooperates with the states in developing the standard forms for data collection as well as standard procedures for data preparation and processing in order to promote a uniform national data base. The degree of uniformity necessary for national statistics has been achieved by periodic issuance of recommended standards by the responsible national agency (NCHS) and the cooperative adoption of these standards by the individual registration areas. These standards include the U.S. Standard Certificates of Live Birth. The standard certificates have been the principal means for achieving the uniformity in information on which national statistics are based; the standard certificate of live birth has been revised 11 times since 1900.

The standard certificates represent the minimum basic data set necessary for the collection and publication of comparable national, state, and local vital statistics data. Also participating in this effort to produce quality, consistent data is the National Association for Public Health Statistics and Information Systems (NAPHSIS), the association of State registrars. Because the State certificates have multiple uses, many factors must be considered and evaluated in deciding what should be included in the recommended standards. Faced with the many uses -- legal, administrative, public health, statistical -- of vital records, NCHS and the state vital statistics offices must make choices regarding the inclusion or exclusion of data items for each revision of the certificate. The NCHS shares in the costs incurred by the states through contractual agreements with each state. Under this arrangement, NCHS obtains and publishes vital statistics based on all births and deaths (e.g., 3,891,494 and 2,314,690, respectively, in 1996) occurring in the United States.

Implementation of the 1997 standards on vital records will require changes in data collection and processing systems at all levels of government and very likely will take at least several years to accomplish throughout the United States. In addition to revising computer systems at the state and Federal levels, the electronic software that is used in hospitals to record and report over 90 percent of all births in the United States needs to be converted. Most importantly, the procedures used to collect birth and death data in hospitals and funeral homes will need to be revised and the appropriate staff will need to be trained.

During 1998 and 1999, NCHS sponsored a committee of state vital statistics officials and representatives of the relevant professions in a series of meetings to evaluate the entire content and format of the current Standard Certificates. It can be anticipated that not all registration areas will implement the 1997 standards at the same time or with complete coverage and compliance at the start. Only one state has implemented the revised race question on birth and death certificates in the year 2000. Most others will wait until the next revisions of the U.S. Standard Certificates of Birth and Death are implemented in 2003.

International Migration Components. As discussed above, the international migration components are based on a variety of administrative sources and analytic estimates. The Immigration and Naturalization Service (INS) supplies data on legal immigrants. The Office of Refugee Resettlement (ORR) supplies data on persons admitted to the United States as refugees. Both sources supply data on country of birth.

To develop data on the race and Hispanic origin of the entering immigrants, the Census Bureau combines the information on country of birth from the INS files with information from the most recent decennial census. Because the INS and other data sources on international migration do not code race or Hispanic origin, no change in these sources is anticipated. The Census Bureau will need to examine the results of Census 2000 and develop new algorithms to accommodate the 1997 categories for data on race.

Internal Migration Components. To develop the internal migration component, the Census Bureau currently relies upon the annual extract of tax returns provided by the Internal Revenue Service (IRS), and a 20 percent sample of information on the Social Security Administration (SSA) Application File (NUMIDENT). Under an agreement between the Census Bureau and SSA, the Census Bureau has recently gained access to a full 100 percent NUMIDENT file. This opens additional opportunities for developing population estimates below the national level by age, sex, race, and Hispanic origin.

This component also presents the biggest obstacle to modifying categories for data on race in the intercensal population estimates process. Under the Social Security system, data on race are provided as part of the Social Security card application process. For the oldest among the population currently covered in the NUMIDENT files, the last application date could refer to the beginning of the Social Security system.

Until 1980, the SSA application system provided three racial categories -- White, Black, and Other. Beginning in 1980, the SSA modified the racial categories on the SSA application form to include five categories -- (1) Asian, Asian-American or Pacific Islander; (2) Hispanic; (3) Black (non-Hispanic); (4) North American Indian or Alaskan Native; (5) White (non-Hispanic). Although SSA modified the racial categories application card, people who already had an SSA card did not have to resubmit their data on race. Thus, pre-1980 entries on the SSA file have information for three racial categories (White, Black, and Other), while entries after 1980 have information for five racial categories. The application for a Social Security card needs to be updated to reflect the 1997 standards.

Another change to the Social Security application procedure has presented challenges to the use of data on race. Beginning in the late 1980's, the SSA introduced the "enumeration at birth program." Under this program, parents could request a social security number for their newborn children with the birth registration process. Because the birth certificates do not include racial information for the newborn, it is impossible to code race for the newborn onto the SSA file. While information on race is available for the birth mother and father on the basic birth registration certificate, these data are not made available to SSA and are not on the basic NUMIDENT file received by the Census Bureau.

5. Comparability Issues

Even the availability of the required source data does not ensure the capability to produce reasonable and accurate population estimates. Production of population estimates by the major demographic characteristics depends upon the availability of comparable data across the various data sources. While comparability issues with respect to race reporting are not new, the increased complexities of new racial categories are likely to exacerbate the problems.

The issues about comparability in race reporting are present in the current set of intercensal population estimates. Data from the 1990 census on race posed several of these problems.

Although the enumeration of the resident population in the 1990 census, without adjustment for net under coverage, was adopted as a standard for the estimates, changes were made to that distribution of the population by age and race. These modifications were made to bring the definition of age and race into conformity with definitions used for data from other sources, such as vital statistics.

For age, the aim was to correct biases in census age tabulations that resulted from displacement of age reporting from the reference date of the census. In 1990 census publications, age is based on respondents' direct reports of age at last birthday, with some editing for age misstatement. This definition proved inadequate for postcensal estimates; many respondents reported their age (even if correctly) at the time of completion of the census form or interview by an enumerator, either of which could have occurred several months after the April 1 reference date. As a result, age was slightly biased upward. Modification was based on a respecification of age, for most individual respondents, according to their year of birth. Age was derived from year of birth by allocating date of birth to the first quarter and last three quarters of each year, subtracting year of birth from 1990 for those born before April 1, and from 1989 for those born after April 1. The allocation was based on an historical series of registered births by month.

For race, the objective of the modification was to conform to the definition of race specified in the 1977 standards. In the 1990 census, a substantial number of people (roughly 9.8 million) did not specify a racial group that could be classified in any of the categories on the census form: White; Black; American Indian, Eskimo, or Aleut; Asian or Pacific Islander. A large majority of these people were of Hispanic origin (based on their response to a separate, Hispanic origin question on the form), and many wrote in their Hispanic origin, or Hispanic origin type (for example, Mexican or Puerto Rican) as their race. People of unspecified race were allocated to one of the four tabulated racial groups (White; Black; American Indian, Eskimo or Aleut; and Asian or Pacific Islander) based on their response to the Hispanic origin question. These four categories for race conform with the 1977 standards, and are more consistent with the categories in other administrative sources than are the original census tabulations.

When combining across data sets and agencies, the problems of comparability in reporting of race become more severe. Clearly, the added complexity of reporting more than one race will add to this problem, particularly as different reporting situations (such as the census or the birth and death certificates) engender differential tendencies to report more than one race. Differences in allocation and editing procedures will almost certainly exacerbate the problem as exemplified by the problem of using data from different data universes in the calculation of rates.

Comparability of race data from the Census 2000 and from vital statistics will be more problematic until the Vital Statistics System adopts the new data collection standards in 2003. NCHS, Census Bureau, and OMB jointly sponsored a workshop in July 2000 which was conducted by the Epidemiology and Biostatistics Department of The George Washington University School of Public Health and Health Services and the Metropolitan Washington Public Health Assessment Center. The aim of the workshop was to develop a research agenda designed to address these problems.

Workshop participants identified several research strategies that could be used to address the technical and methodological issues that the 1997 standards for data on race and ethnicity raise for vital statistics and population estimates, and to address some of the more conceptual and perceptual aspects of race and ethnicity that influence reporting in censuses, surveys, and vital records. A summary of the workshop is being prepared. Selected research strategies are summarized below:

> *Census or Survey-based Matching Studies:* In such studies, the race and ethnicity reported on birth or death certificates could be compared with the race and ethnicity reported for those individuals in the decennial census or surveys such as the American Community Survey, the Current Population Survey, and the National Health Interview Survey. The participants also recommended the continuation of the National Longitudinal Mortality Studies.

> *Studies of Selected Vital Records:* Although NCHS works with the states to establish a standard set of data on births and deaths to be reported to the Federal Government, the original data collection process is established and managed by each individual state. As a result, some states are collecting additional data on race and ethnicity that could be used to study features of race and ethnicity reporting in vital records and to explore how reporting may change under the 1997 standards. NCHS is investigating the possibility of obtaining additional race information for a selected number of states for methodological purposes.

> *Studies to Identify Sources of Error:* A careful study of the potential sources of error -- both bias and random variation -- in the data and procedures used to produce population estimates would provide valuable guidance to both the producers and users of those estimates. Some of the sources of error include the census undercount and undercount adjustments, missing data and imputation procedures, and misreporting and misclassification of data. The 1997 standards will not change the types of error in the data but will introduce specific new forms of error, such as misreporting or misclassification between single- and multiple-race categories, that need to be

80

understood. The proposed matching studies would help identify and quantify some types of bias in the data. Techniques like multiple imputation can be used to assess variance.

Studies of the accuracy of estimates and projections: The Census Bureau produces population estimates and projections for states and communities, but states also produce independent estimates of their populations, which may differ from Census Bureau estimates. Studies are needed to compare the accuracy of Census Bureau and state estimates and to identify the sources of discrepancy.

Studies on Race and Ethnicity: Proposals for other studies emphasized opportunities to improve data on race and ethnicity by gaining a better understanding of the meaning of racial and ethnic identities in society and of the factors that influence both self-identification and identification of others. One suggestion was to look at how race and ethnicity were reported in Census 2000 for those in multiple race households. Studies might also use routinely conducted surveys like the Current Population Survey or the National Health Interview Survey to ask about the race of respondents' parents and other ancestors to learn more about the older roots of multiple race identities.

An additional source of noncomparability results from the fact that the standard certificate of live birth includes items asking for the race of the mother and the race of the father. At no time have the certificates included a question on the race of the child. In 1988 and prior years, NCHS assigned a race to the child solely for statistical purposes. Births were tabulated by this assigned race of the child, which was created from information reported for the race(s) of the parents as entered on the birth certificate. When the parents were of the same race, as was the case for 95 percent of births with race reported for both mother and father in 1998, the child was assumed to be of the same race. If the parents were of different races and one parent was white, the child was assigned the race of the parent who was not white. When the parents were of different races, and neither parent was white, the child was assigned, for statistical purposes, to the father's race, with one exception: if either parent was Hawaiian, the child was assigned to Hawaiian. If race was missing for one parent, the child was assigned the race of the parent for whom race was reported.

In 1989, NCHS changed its tabulation procedures and began *tabulating* births by the race of the mother. The most important factor influencing this decision was the 1989 revision of the birth certificate. This revision includes many more health questions that are directly associated with the mother (e.g., method of delivery, medical risk factors, tobacco and alcohol use, and maternal weight gain). Many of the other items that have been on the birth certificate since 1969 or even earlier also relate directly to the mother; for example, age, educational level, and receipt of prenatal care. In all these instances it is more appropriate to tabulate births by the mother's race.

A second factor influencing the decision to *tabulate* births by race of mother is the large proportion of births with race of father not reported, 14 percent in 1998. Although this proportion has declined slightly in the 1990's, it is still higher than in 1978, 11 percent. The high proportion of records with the father's race not reported reflects the increase in the proportion of births to unmarried women; in many such cases, no information is reported on the father. These births are already assigned the race of the mother because there is no alternative. Tabulating all births by race of mother, therefore, provides for a more uniform approach, rather than a necessarily arbitrary combination of parental races.

It is important to keep in mind that the public use data files that NCHS produces and disseminates include both the mother's and father's races as reported on the birth certificate. Researchers and others may *tabulate* the birth data by race of the mother, by race of the father, or by race of the child as determined by the algorithm previously used by NCHS, or an alternative as they wish. For purposes of research based on data from the birth certificate itself, it is generally most appropriate to tabulate the birth data by the race of the mother. NCHS has re-tabulated all of the trend data on births by race of mother for the years beginning with the 1980 data year. To facilitate continuity and analysis of the data, trend tables showing data for years prior to 1980 show data for both race of mother and race of child. This makes it possible to distinguish the effects of this change from real changes in the data.

There continues to be interest in collecting and reporting birth data according to the "race of the child" and this would be particularly relevant for developing population estimates. Over the last several years, this concept has been discussed by the panel charged with developing the forthcoming revision of the U.S. Standard Certificate of Live Birth. The panel revising the certificate did not recommend that an item on the child's race be added to the birth certificate. As mentioned earlier, each state has its own certificate of live birth, and in most cases state certificates closely follow the content and format of the U.S. Standard Certificate. Individual states may choose to add an item asking for the "race of the child," even though the item is not on the U.S. Standard Certificate of Live Birth. NCHS and colleagues in state vital statistics offices continue to research this issue.

6. Future Direction

The process of developing a set of intercensal population estimates consistent with the 1997 standards will not be an easy one. Until data are available, making any commitments about the probable set of products is impossible. The Census Bureau realizes, however, that many data users need to know its plans in order to make their own program decisions.

To begin this process, the Census Bureau is forming a technical interagency group of key data providers and key data users to address many of the major issues. Members of this group will

provide input on: (1) the feasibility of using one consistent set of categories on race across all geographic levels; (2) the feasibility of using population size as the only criterion for determining which categories by race will have separate population estimates; (3) the minimum cell size below which population estimates will not be produced; (4) the continued development of population estimates by mutually exclusive categories on race; and (5) the use of consistent methodologies for the different categories by race in the population estimates program. This technical group will also examine issues related to data allocation and editing -- important factors related to the data consistency issues.

Detailed data on race from Census 2000 will not be available until mid 2001. In the meantime, the interagency group can address and reach consensus on most of the issues outlined above. Through these discussions with the data providers and data users, the Intercensal Population Estimates Program can begin to form some tentative plans. Although it is too soon to speculate on any outcomes, it is likely that the Intercensal Population Estimates Program will need to be flexible. During the coming decade, as more data become available using the 1997 standards, it is likely that the Census Bureau will continue the expansion of the population estimates program to include additional categories by race.

C. Uniform Crime Reporting Program

The Federal Bureau of Investigation's (FBI) Uniform Crime Reporting (UCR) Program was established by the International Association of Chiefs of Police in 1929. The UCR Program has as its primary mission the collection of uniform national crime statistics for the express purpose of assisting law enforcement across the Nation in strategic planning in the war against crime. Currently, nearly 17,000 city, county, and state law enforcement agencies voluntarily participate in the program.

Even though it is often assumed that UCR statistics consist of a monolithic statistical collection effort, the UCR Program in reality comprises several programs, all with different data collection requirements. Approximately 89 percent of the crime reported to the FBI is through the Summary Reporting System. The Summary System reflects aggregate tallies of crime information from law enforcement agencies, and the collection of data on race is limited to arrest information only. UCR data are additionally collected through the Law Enforcement Officers Killed and Assaulted Program, the Hate Crime Data Collection Program, and the National Incident-Based Reporting System – all of which provide for the collection of data on race. Race is usually determined by observation on the part of the responding officer or the victim of the crime, rather than self-identification. Any effort to collect information on race by self-identification is either difficult or impossible due to the reliance on third-party information or the sensitive nature of handling the response to a criminal incident.

The information collected in the UCR Program is a complex derivative of law enforcement operational databases. As such, the primary application of any UCR statistics, including those involving race, is by law enforcement in the examination of crime trends. The data provide law enforcement a unique means by which to conduct analyses, ranging from limited agency-specific to broadly multijurisdictional, in order to plan strategically for the maximum use of limited law enforcement resources.

In concurrence with other national criminal justice operational databases managed by the FBI (e.g., National Crime Information Center and the Integrated Automated Fingerprint Identification System, etc.), the UCR Program currently uses the four race categories specified by the 1977 standards. Incorporation of the 1997 standards into the UCR Program is being considered within the context of the burden such implementation would place upon law enforcement agencies in their participation in a voluntary program, and the feasibility of collecting the information in accordance with the 1997 standards within the confines of the normal law enforcement operational setting.

CHAPTER 5

COMPARING DATA COLLECTED UNDER THE 1997 AND THE 1977 STANDARDS

This chapter is a summary of the Bridge Report: Tabulation Options for Trend Analysis, which is provided in Appendix C.

A. Introduction

Agencies whose data are used to display time trends in economic, social, and health characteristics by racial and ethnic groups may need to consider bridging methods to assist users in understanding the data collected under the 1997 standards. For some period of time, referred to as the bridge period, agencies may use two estimates. The first, a tabulation of the data collected under the 1997 standards (see Chapter 3, Part B), and the second, a "bridging estimate" or prediction of how the responses would have been collected and coded under the 1977 standards. The bridging estimate is designed only for analyzing historical trends in data series. Once the bridge period is over, the bridge estimates will no longer be needed.

It should not be assumed that bridging is useful or required in every situation. Agencies should carefully consider whether they need bridging estimates. Bridging estimates may not be needed if agencies can tolerate a "break" in their data series or if comparison to another data series provides users with enough information about the change. If bridging estimates are not used, however, agencies should footnote the first occurrence of data collected under the 1997 standards.

There are at least two purposes of bridge estimates: (1) to help users understand the relationship between the old and new data series (as noted above); and (2) to provide consistent numerators and denominators for the transition period, before all data are available in the new format. If there is a need for bridging, agencies should carefully evaluate alternative methods. The work presented in Appendix C, and summarized below, is intended to help inform agencies about the statistical characteristics of selected bridging methods.

Agencies are encouraged to plan and conduct methodological research that will lead to more informed decisions concerning bridging methods and their uses. Such methodological research has long been used to quantify changes in data collection procedures. For example, when methods for coding industry, occupation, or diseases are updated, it is common practice to code data using both sets of coding rules to determine the nature and extent of the changes introduced by the change in procedures.

The analyses presented in Appendix C make use of survey data in which the same respondent provided racial information in response to both a question structured under the 1977 standards and to questions similar to those that might be structured under the 1997 standards. These are examples of methodological approaches that can be adopted by agencies, if necessary. In particular, since 1976, the National Health Interview Survey (NHIS) has added a follow-up question for those reporting more than one racial identity, asking them to select the one that they feel best describes them. This information is directly used in some of the most promising bridge techniques. Some agencies may want to use such a follow-up question experimentally to provide valuable, survey-specific information for bridging to the past. As agencies conduct such experiments, the results may assist other agencies in understanding the changes associated with transitioning to the 1997 standards.

The results discussed here and in Appendix C represent the work of a group of statistical and policy analysts drawn from Federal statistical agencies that use and produce data on race and ethnicity. They have spent the past three years considering these tabulation issues and conducting research to develop tabulation guidelines for constructing "bridges" between racial data collected under the 1997 standards and racial data collected under the 1977 standards. The report sets forth criteria by which different bridging methods should be evaluated and describes the different methods that have been considered thus far. The results of the research conducted on several methods for creating bridges are also presented.

This part of the report discusses different options for tabulating racial data in order to create bridges from data collected under the 1997 standards, which have five racial categories and permit the reporting of more than one race, back to the data collected under the previous standards, which identified four racial categories and asked respondents to select one. An "Other" category appears in much of the analysis, because it is included in the decennial census and some other surveys.

All of these methods (and the research on them reported here) involve the use of individual-level records. Analysis is limited to data collected using the separate questions for race and Hispanic origin. Under the 1997 standards, when reporting is based on self-identification, the two-question format is to be used; even in the case of observer identification, this is the preferred format. It is expected that some users will bridge to a distribution created using the combined format for the question on race and ethnicity. Thus, bridging both to the old racial distribution arising from the use of two questions and one based on a combined, single question are analyzed. The latter analyses required the creation of a combined distribution from data collected using the two-question format. It should be clearly understood that this is a "manufactured" distribution and may be different from one obtained when actually using a combined question format. Based on the research, the strengths and weaknesses of each bridging method are discussed. The last two sections of this chapter discuss weighting data collected under the

1997 standards to 1990-based population controls and suggested bridging strategies for users of the new racial and ethnic data.

B. Methods for Bridging

The goal of developing bridging methodology for data on race is to identify a statistical model that will take individuals' responses to the new questions on race and classify those responses as closely as possible to the responses we hypothesize they would have given using the old single race categories. Such a task will be relatively easy or be more difficult depending on how an individual identifies himself or herself under the 1997 standards. For bridging purposes, individuals with a single racial background are likely to identify as they did before, and no statistical model is needed for bridging. However, those with a mixed racial heritage who were previously required to identify only one part of their background may, under the 1997 standards, choose to report more than one racial identity. When a person identifies with more than one racial group, some model will be necessary to translate those multiple responses into the one, single response we hypothesize that the individual most likely would have reported under the 1977 standards.

1. Framework

Several different methods have been identified for creating a single race distribution from data including multiple race responses. These methods vary in both the assumptions that are made and the procedures that are followed. Before describing the particular methods examined in this report, it is useful to describe some of their major underlying characteristics.

One major distinction among the methods is whether an individual's responses are assigned to a single racial category (termed whole assignment) or to multiple categories (termed fractional assignment). Whole assignment can be based on a set of deterministic rules or based on some probabilistic distribution. For example, a deterministic rule might assign all White and American Indian responses into the American Indian category, while a probabilistic rule might randomly assign 60 percent of the White and American Indian responses into the American Indian category, and 40 percent into the White category. In the above example, it is unlikely that all individuals identifying as White and American Indian under the 1997 standards would have previously identified as American Indian, so the deterministic rule will result in misclassifications for all those people who had previously identified as White. With a probabilistic rule, an individual's responses are randomly assigned to either the American Indian category or the White category (such as with 60 percent and 40 percent probabilities, respectively, based on previously collected data). However, even if the overall probabilities matched exactly the aggregate distribution under the 1977 standards, there is no guarantee that the 40 percent who were categorized as White would have classified themselves that way. In fact, in the worst case,

all 40 percent who were classified as White would actually have identified as American Indian under the 1977 standards, and a corresponding percentage of those categorized as American Indian would have identified as White.

When fractional assignment is used, multiple race responses are categorized into more than one category where each category receives a fraction of a count, and the sum of the fractions equals one. In the above examples of whole assignment, a person's responses were placed into one and only one category, in an attempt to mimic the past. An alternative is to use a deterministic rule to assign some fraction of the multiple race responses to each of the racial categories identified. For example, a multiple response of White and American Indian might count as "one-half" in the tabulations for American Indians and "one-half" in the tabulations for Whites. These fractions, like the probabilities in the earlier example, could be varied for different combinations of multiple races to attempt to reflect how often people might identify with one group compared with another.

2. Bridge Tabulation Methods

All of the bridge tabulation methods focus on the assignment of the responses from individuals who identify with more than one racial group. Responses from individuals who identify with a single racial group under the 1997 standards are assumed to have been the same under the 1977 standards. The response "Native Hawaiian or Pacific Islander" is assigned to the old racial category of "Asian or Pacific Islander." The specific methods for assigning multiple race responses into single race categories are Deterministic Whole Assignment, Deterministic Fractional Assignment, and Probabilistic Whole Assignment.

Two sets of results for each of the following tabulation methods are produced. The first set ignores the use of any auxiliary information other than that needed to carry out the particular tabulation method. The other set of results for each method uses the one piece of information that is certain to be common to all data collections done following the 1997 standards, that is, ethnicity. Thus, whether or not an individual is Hispanic is taken into account when a tabulation method is used.

Deterministic whole assignment. This method uses fixed, deterministic rules for assigning multiple responses back to one and only one of the racial categories from the 1977 standards. Four alternatives are examined. The first (Smallest Group) assigns responses that include White and another group to the other group, but responses with two or more racial groups other than White are assigned into the group with the fewest number of individuals identifying that group as a single race. The second alternative (Largest Group Other Than White) assigns responses that include White with some other racial group, to the other group, but responses with two or more racial groups other than White are assigned into the group with the highest single-race

count. The third alternative (Largest Group) assigns responses with two or more racial groups into the group with the largest number of individuals as a single race. In this latter case, any combination with White is assigned to the White category, and combinations that do not include White are assigned to the group with the largest single-race count. The fourth alternative (Plurality) assigns responses based on data from the National Health Interview Survey (NHIS). The NHIS has permitted respondents to select more than one race for a number of years, with only the first two responses captured. However, respondents reporting more than one race were given a follow-up question asking them to select the one race with which they most closely identify (called Main Race here). For these respondents, the proportion choosing each of the two possibilities as their main race was calculated. All responses in a particular multiple-race category using the Plurality method are assigned to the group with the highest proportion of responses on the follow-up question about main race.

Deterministic fractional assignment. This method uses fixed, deterministic rules for fractional weighting of multiple-race responses, that is, assigning a fraction to each one of the individual racial categories that are identified. These fractions must sum to 1. Two alternatives are examined. The first (Deterministic Equal Fractions) assigns each of the multiple responses in equal fractions to each racial group identified. Thus, responses with two racial groups are assigned half to each group; those with three groups are assigned one-third to each, etc. The second alternative (Deterministic NHIS Fractions) assigns responses by fractions to each racial group identified, with the fractions drawn from empirical results from the NHIS (as described above).

Probabilistic whole assignment. This method uses probabilistic rules for assigning multiple race responses back to one and only one of the previous racial categories. Two alternatives are examined. These parallel the two alternatives discussed under Deterministic Fractional Assignment, except that, for a given set of fractions, the response is assigned to only one racial category. The fractions specify the probabilities used to select a particular category. The first alternative uses equal selection probabilities. The second uses the NHIS fractions where possible, and equal fractions when no information is available from NHIS. Probabilistic Whole Assignment will yield nearly, on average, the same population counts as Deterministic Fractional Assignment.

Only the results from Deterministic Fractional Assignment are presented in this report. In practice, there would be a difference between Deterministic Fractional Assignment and Probabilistic Whole Assignment when computing variances for tabulated estimates, and the two methods would yield relatively small differences in distributions for respondent characteristics. In general, Probabilistic Whole Assignment would yield a higher estimated variance than the Deterministic Fractional approach, with the variances for both methods underestimating the true variance. Probabilistic methods which incorporate a "Multiple Imputation" statistical technique

would result in an unbiased estimate of variance, but at the price of being more difficult to implement (See Rubin 1987.).

All Inclusive. A final tabulation method considered is termed the "All Inclusive" method. Under this method all responses are used. Responses are assigned to each of the categories that an individual selects. The sum of the categories totals more than 100 percent.

C. Methods of Evaluation

1. Data Sources

National Health Interview Survey. The NHIS is a continuing nationwide sample survey designed to measure the health status of residents of the United States (Benson and Marano, 1995; Massey et al., 1989). The analysis here uses data from an analytic file that contains three years of NHIS data (1993, 1994, and 1995). For each of these years there were about 45,000 households interviewed, resulting in slightly more than 100,000 individuals per year. The total sample for the bridge analysis is 323,080 (5237 respondents did not provide data on race). Since 1976, the NHIS has allowed respondents to choose more than one racial category. As the respondent is handed a card with numbered racial categories, the interviewer asks, "What is the number of the group or groups that represents your race?" If a respondent selects more than one category, the interviewer then asks, "Which of those groups would you say best describes your race?"

Although the listed racial groups have changed over time, for 1993 to 1995, the card shown to respondents included 16 separate racial categories (White, Black, American Indian, Aleut, Eskimo, Chinese, Filipino, Hawaiian, Korean, Vietnamese, Japanese, Asian Indian, Samoan, Guamanian, and other Asian and Pacific Islander. In addition, although not on the flashcard, respondents were allowed to give an "other" race response.) To be consistent, the 16 groups were collapsed to the four previous racial categories: White, Black, American Indian or Alaskan Native (AIAN), and Asian or Pacific Islander (API), plus Other.

For this analysis, a variable called Detailed Race was created from responses to the first question, which allowed identification with more than one racial group. This information is not included on public use data files of the NHIS. However, on internal files, the first two race groups mentioned are recorded for each observation. Even if a respondent selected more than two groups, only two were recorded on the intermediate file. From the two recorded racial responses, Detailed Race was coded into five single race groups (White, Black, AIAN, API, Other) and 11 multiple race groups (White/Black, White/AIAN, White/API, White/Other, Black/AIAN, Black/API, Black/Other, AIAN/API, AIAN/Other, and API/Other. For most analyses, multiple race combinations that had insufficient numbers were aggregated into the

category "Other Combinations.") Individuals who had two racial groups recorded for Detailed Race but a third group recorded for the "group that best describes race" were coded into "Other Combinations."

The Main Race variable, used as a reference point representing the racial distribution under the 1977 standards, is primarily derived from Detailed Race and the responses to the second question, which asks the respondent for the group that best describes his/her race (Benson and Marano, 1995). For respondents who selected one Detailed Race group, Main Race is the same as Detailed Race. For respondents who selected more than one racial group, Main Race is the one group reported as best describing their race. Some respondents who had chosen more than one race for the Detailed Race question responded as "Multiple race" or "Other" for the Main Race question. For this analysis, these responses were combined into the "Other" category. Categories for Main Race were White, Black, AIAN, API, and Other.

The combined race and ethnicity variable, referred to here as "Combined Main Race," uses the respondent's answer to an Hispanic origin question to reassign the respondent to the Hispanic category. The Hispanic origin question used is the following: Are any of these groups (the respondent's) national origin or ancestry? At the same time, the interviewer hands the respondent a card listing Hispanic groups as categories: Puerto Rican, Mexican, Cuban, Mexican American, Chicano, Other Latin American, and Other Spanish. For this report, Whites, Blacks, Others, and those reporting more than one race who identified with any of the Hispanic groups were categorized as Hispanic and not according to their race. Asians, American Indians, and Alaska Natives were not reclassified. If a respondent did not answer the Hispanic origin question, he or she was assumed to be non-Hispanic.

May 1995 Supplement on Race and Ethnicity to the Current Population Survey (CPS). The May 1995 CPS Supplement was one in a series of studies conducted for the Federal agencies' review of the standards for data on race and ethnicity. The Supplement was designed to address the following issues: (1) the effect of having a "multiracial" race category among the list of races; (2) the effect of adding "Hispanic" to the list of racial categories; and (3) the preferences for alternative names for racial and ethnic categories (e.g., African-American for Black, and Latino for Hispanic). The Supplement was organized into four panels representing a two-by-two experimental design for studying the first and second issues outlined above. Each panel was given to one-fourth of the sample, or about 15,000 households (30,000 individuals). All respondents in a household received the same set of questions; household members 15 years and older were asked to respond for themselves, and parents answered for children under 15.

Only two of the panels in the CPS Supplement permitted respondents to report in a multiracial category (panels 2 and 4), and only one panel had separate race and Hispanic origin questions (panel 2) as ultimately recommended in the 1997 standards. Therefore, panel 2 data were used

to analyze the effects of the different tabulation methods for the two-question format. The new Hispanic question, preceding the race question, simply asked whether or not the respondent was Hispanic. The smaller sample (about 30,000 observations) hampers analysis and generalizations when the focus is on the small portion of the sample (about 1 percent) who identified as "multiracial."

There are additional limitations to these data for evaluating the bridging methods. The option respondents were given to identify multiple races in the CPS Supplement was a multiracial category with a follow-up question asking respondents to indicate all the racial groups with which they identified. The 1997 standards allow people to identify directly with all the racial groups they choose and do not include a "multiracial" category. Furthermore, a large percentage of individuals who chose the multiracial category in panel 2 of the Supplement did not specify more than one racial group (see Tucker et al., 1996). For purposes of this evaluation, individuals were classified as belonging to the specific racial categories they identified. Those who identified as being multiracial but then did not give two or more specific racial groups were reclassified in the one racial category they gave. *Thus, the distribution of the CPS Supplement data reported here differs from that which was published in earlier reports, which classified as multiracial any person who identified with the multiracial category even if they only specified one racial group.* This new distribution is referred to here as the "Edited Distribution."

This edited distribution was used with the various tabulation methods. As in NHIS, the resulting distributions were compared to a reference distribution based on the respondents' original answers (in the first CPS interview) to the race question that followed the 1977 standards.

The combined race and ethnicity format, still referred to as the reference distribution in the relevant tables, uses the respondent's answer to the Hispanic origin question to reassign the respondent to the Hispanic category. For this report, Whites, Blacks, Others, and those reporting more than one race who identified with any of the Hispanic groups were categorized as Hispanic and not according to their race. Asians, American Indians, and Alaska Natives were not reclassified. If a respondent did not answer the Hispanic origin question, he or she was assumed to be non-Hispanic.

1998 Washington State Population Survey. The 1998 Washington State Population Survey (WSPS) was designed to provide information on Washington residents between decennial censuses. The survey collected data on employment, income, education, and health, along with basic demographic information. The WSPS was done by telephone and included 7,279 households with telephones. Blacks, Asians, Hispanics and American Indians were over sampled. The designated respondent was the individual with the greatest knowledge about the household. The respondent weights reflect this over sampling and, thus, results are representa-

tive of the Washington population as a whole. The response rate for the entire sample was between 50 and 60 percent.

Information about the race of the respondent was collected twice during the course of the interview. At the beginning of the survey, the respondent was asked, "Are you of Hispanic origin?" Following that question, the respondent was asked, "What is your race?" The categories were the ones appearing under the 1977 standards, but the order was as follows: Black; American Indian, Aleut, or Eskimo; Asian or Pacific Islander; and White. An "Other" category also was allowed, and the interviewer recorded the verbatim response on a "specify" line. Near the end of the survey, the respondent was asked race questions conforming to the 1997 standards. Besides the same Hispanic origin question, the respondent was asked to specify country of origin. For race, the respondent was asked to select one or more categories. This time the ordering of the categories was White; Black or African American (or Haitian or Negro); American Indian or Alaska Native; Native Hawaiian or Other Pacific Islander; Asian. Again, an "Other" category was provided. There also was a follow-up question for Asian respondents to specify country of origin. The results from the race question at the end of the survey were used with the tabulation methods. The reference distribution came from the answers to the original race question. A combined race/ethnicity format was not created from the WSPS data because unedited information from the race question using the 1977 format was unavailable.

2. Advantages and Disadvantages of These Data Sources

Only the Washington State data closely resemble the way the question on race will be asked under the 1997 standards. Yet, all three can offer insights into the relationship between how individuals will actually respond to the 1997 question on race and how they responded to the question under the 1977 standards. The NHIS and the CPS Supplement are nationally representative, and the Washington State data serve as an example for evaluating the tabulation methods at the state level. Simulations using 1990 census data also were conducted, but the results differed little from those for the other data sets. At this point, it is believed that an analysis of data from the 1998 Dress Rehearsal for Census 2000 would be of greater utility. Furthermore, the Dress Rehearsal data will provide examples of the effects of the 1997 standards at the local level. Thus, this analysis will be included in the next version of this report.

3. Description of New Analyses

The analyses concentrated on the bridge tabulation methods. These analyses can be divided into three broad areas: (1) descriptions of racial distributions under the alternative bridging tabulation methods; (2) rates of racial "misclassification" for these alternatives; and, (3) sensitivity of outcome measures to the bridging alternatives.

Distribution of Race. For the first phase of the analysis (using the NHIS, the CPS Supplement, and the data from Washington State), the distributions of race under the allocation alternatives described previously were calculated: All Inclusive, Deterministic Whole Allocation (Smallest Group, Largest Group Other Than White, Largest Group, and Plurality) and Fractional Allocation (Equal Fractions and NHIS Fractions). These new distributions were compared to the reference distribution in each data set. At this time, it is unknown what percentage of people in the United States will identify with more than one racial group when given the opportunity to do so in Census 2000 and in subsequent surveys. For purposes of illustrating the effects of a greater proportion of individuals identifying more than one racial background, analyses were conducted increasing the proportion of multiple race responses two-, four-, six- and eight-fold using the NHIS, the CPS Supplement, and the Washington State micro data sources. The racial distributions were compared using each of the tabulation methods to see effects with increasing levels of reporting more than one race. Of necessity, these tabulations assume that the increases are the same across the different combinations of more than one race. The accuracy of this assumption cannot be tested. The purpose of these analyses is not to attempt to make accurate predictions about the extent of multiple race reporting or its composition, but rather to see more clearly possible differences among tabulation methods that may become apparent only with a greater percentage of more than one race reporting.

Misclassification of Race. Besides evaluating the overall racial distributions produced by the tabulation methods, the misclassification of individuals also needs to be examined. For the NHIS, the CPS Supplement, and the Washington State survey, these misclassification rates were formed by comparing an individual's answer to the race question under the 1977 standards to the assigned category of the individual's response(s) to the race question under the 1997 standards using each of the tabulation methods. The misclassification rate and its standard error for each race by tabulation method were produced.

Preliminary Outcomes Assessment. In the last phase of the analysis, the impact of multiple-race reporting on outcome measures was assessed. This process is important because users in many of the Federal agencies are not typically examining race distributions, but rather trends and indicators for the Nation (e.g., health outcomes, economic well-being, educational attainment) across racial groups. This is where the majority of work will need to be done within individual agencies as the 1997 standards are implemented. An initial examination of how common statistics could be affected by reporting of more than one race was conducted. Five outcome measures were examined, three from the NHIS and two from the CPS Supplement. From the NHIS, three routine health outcomes were calculated: percent of respondents in poor or fair health, percent of children living with a single mother, and percent of respondents with no health insurance. From the CPS Supplement, the proportions of respondents who were unemployed and the labor force participation rates for different racial groups were calculated. These estimates based on the bridging alternatives are not meant to be precise measures of these

factors, but are used to demonstrate the possible impact reporting of multiple races and the tabulation methods may have on these and similar estimates.

D. Examination of the Results with Respect to the Evaluation Criteria

Bridging to the past will be needed for measuring change in a variety of circumstances.   Besides measuring population growth, any number of economic, social, and health outcomes must be monitored. This work will involve different population groups at different levels of geography. As a first step toward providing the information users will need to make informed decisions about the methods, the strengths and weaknesses of the bridging methods with respect to the evaluation criteria outlined at the beginning of this report are discussed, based on the results of the statistical analyses conducted. The details of these statistical analyses can be found in Appendix C.

Measure Change Over Time. As indicated earlier, measuring change over time is the criterion that is of greatest importance in evaluating the bridging methods. The first and second phases of the analysis shed light on the performance of the various methods in this area. In essence, an ideal bridging method in this case is one that not only accurately recreates the population distribution under the 1977 standards such that the only difference remaining is a function of true change over time, but also assigns an individual's response to the old category that would have been chosen. The methodology used in these studies allows users, within limits, to see how well the bridging methods using racial data collected under the 1997 standards can match data from the same respondents collected (at about the same time) under the 1977 standards. To the extent that there is a match, any change that would occur from this point forward would indicate true change. If the match is poor, it is not possible to isolate the true change.

When comparing the different methods to their reference distributions (whether using only the race question or the combined format), the racial categories that were most sensitive to which method is chosen were the numerically small ones, particularly the AIAN category. While different data sets were used in each study and the racial questions were not the same, the studies indicate that the Largest Group Deterministic Whole Assignment method, the Plurality method, and the two Deterministic Fractional Assignment methods produce distributions closer to the reference distributions than do the other Deterministic Whole Assignment methods and the All Inclusive method. Controlling for ethnicity had no effect on these results. One reason the Largest Group Assignment method results are so close is that it has little effect on the smaller racial groups, because most assignments are made to Black or White, and the percentages for these two races are so large that the relatively small increase they receive is not noticeable. The Plurality method produces a good fit, because it makes assignments at the level of specific racial combinations. The performance of the NHIS Fractional Assignment method can be discounted to a degree in the NHIS study because the analysis is somewhat circular; however, the results

from the CPS Supplement and the Washington State Population Survey (WSPS) show this method yields a relatively close match. The Equal Fractional Assignment method produces a reasonable match in these studies. The primary reason that the other two Whole Assignment methods and the All Inclusive method do not perform as well is that they alter the White percentage to some extent and substantially increase the percentage in the AIAN category.

In the case of misclassification rates, some contradictory results emerge, both when using the race question alone and when using the combined format. While the AIAN and "Other" categories have high misclassification rates across all tabulation methods in the CPS Supplement, the same is not true for the other two surveys. The Smallest Group Whole Assignment and the Largest Group Other Than White Whole Assignment methods produce the most comparable results for the AIAN category in both surveys and for the "Other" category in the WSPS; however, these methods have higher overall misclassification rates. Both the CPS Supplement and the WSPS have large misclassification rates for these two categories when using many of the tabulation methods.

When the distributions of the outcome variables are examined, all methods produce comparable, and relatively close matches for all health outcomes. For the AIAN unemployment rate, the Largest Group Whole Assignment method and the NHIS Fractional Assignment method appear to produce the least comparable results, but none of the differences is significant. There are significant differences in the AIAN labor force participation rates for several of the tabulation methods. It is likely that which method is best at matching a reference distribution for outcome measures will depend on the outcome being examined. Unfortunately, the data to assess the best tabulation method for each outcome may never be readily available.

All of these conclusions should be viewed with caution. Many assumptions had to be made in these studies. It is unclear how people will respond to the new racial question in the future, and these responses could differ by mode of data collection and with the subject of the survey. Furthermore, most of this work on developing bridging methods relied on sample data, and small samples at that.

Congruence with Respondent's Choice. This criterion concerns how well the full range of the respondent's choices is represented in the racial distribution. It is more important for evaluating ongoing tabulations under the 1997 standards, but the bridging methods can be differentiated with respect to this criterion, too. None of the Deterministic Whole Assignment methods takes into account the full range of the respondent's selections, but the Plurality method at least controls for the particular racial combination chosen by the respondent under the 1997 standards. The All Inclusive method accurately reflects all selections by tabulating actual responses and not people. The Equal Fraction Assignment method tabulates people, but, like the All Inclusive method, treats all responses equally. The NHIS Fractional Assignment method takes all

96

responses into account, but assignment is based on attempting to estimate in which single-race category the respondent would prefer to be counted.

Range of Applicability. This criterion refers to how well the bridging method can be applied in different contexts. The All Inclusive method provides the same results in every context, because assignment does not depend on the particular detailed racial distribution. This method is not suitable for users who need a distribution that adds to 100 percent. Of the Deterministic Whole Assignment methods, the Largest Group Assignment method is the least sensitive to context and can be used in a wide variety of applications. The other Deterministic Whole Assignment methods are as easy to use as the Largest Group Whole Assignment method, but the results for the small racial categories will vary to a greater extent with the context, particularly according to level of geography. The Equal Fraction Assignment method is as generalizable as the All Inclusive method, but it is not quite as easy to use. The NHIS Fractional Assignment method and the Plurality method may be the most problematic, because they currently represent only a national preference distribution based on data from 1993 to 1995. The use of this distribution at the local level would be likely to produce inaccurate results in a number of cases. That is not to say that the other methods do not face the same problem.

Meet Confidentiality and Reliability Standards. Because these methods all attempt to reproduce the racial categories under the 1977 standards, the same confidentiality problems that existed over the last 20 years will continue to exist. No increase in problems is anticipated. In the case of reliability, however, the situation is different. The All Inclusive method will not produce less reliable data than data produced under the 1977 standards. The Equal Fraction Assignment method may have reliability problems as a result of adding only fractional counts to some of the smaller categories if these categories have a high probability of being chosen as the preferred single race. The same would be true if equal fractions were used to make whole assignments. In sample surveys, the Deterministic Whole Assignment methods will have reliability problems to the extent that there is a large variance on the individual race proportions. This is likely to occur when small samples are involved. The Largest Group Whole assignment method should have the fewest problems with respect to reliability, and the Smallest Group Whole Assignment method will likely have the most. These methods have another problem, however, in that an individual's response may be assigned to different categories at different levels of geography. The NHIS Fractional Assignment method, as well as methods where fractions are used for whole assignment (i.e., the Plurality method), is based upon a sample distribution with its own variance properties. Reliability for the very small combinations will be quite bad unless many years of data are combined, and this presents its own problems.

Minimize Disruptions to the Single Race Distributions. This criterion is relevant only for evaluation of bridging methods. Its purpose is to see how different the resulting bridge distribution is from the single-race distribution for detailed race under the 1997 standards. To the extent

that a bridging method can meet the other bridging criteria and still not differ substantially from the single-race proportions in the ongoing distribution, it will have value for looking both forward and backward in time. An evaluation of the different methods according to this criterion involves the comparison of the bridge distributions to the detailed race distribution under the 1997 standards in each case.

For the CPS Supplement, the Plurality method is marginally closer than the Largest Group Whole Assignment method and the Fractional methods. While the All Inclusive method and the other Deterministic Whole Assignment methods match for the White category, they differ substantially from the single-race AIAN category in the detailed distribution and are marginally worse for the API category. The NHIS Fractional method is the closest in both the NHIS and WSPS.

Statistically Defensible. To be statistically defensible, the bridging method must conform to acceptable statistical conventions. The All Inclusive method makes no assumption about how respondents would assign themselves in the single race situation. The NHIS Fractional Assignment method and the Plurality method are based on an observed distribution, and, to that extent, involve less judgment than the rest of the methods that assign people and not responses. While the Equal Fractional Assignment method is based on judgment, it does not make assumptions about the relative importance of any given race. The Largest Group Whole Assignment method does assign greater importance to one of the races, but it also follows common, but different, statistical practice than the equal fraction approach. Both attempt to minimize the error in assignment. The Smallest Group Whole Assignment method and the Largest Group Other Than White Whole Assignment method do not follow statistical practice, but, instead, rely on the historical record of discrimination; even in these cases, however, the assigned category is based on an observed distribution.

Ease of Use. "Ease of use" refers to how complicated it is to produce the bridge results. The Equal Fractional Assignment method makes assignments that do not depend on the particular detailed racial distribution at hand. It and the NHIS Fractional Assignment method do require the duplication of individual records or the creation, on every record, of a variable for each racial category under the 1977 standards in order to be able to assign fractions for any combination of categories. If the fractional methods are used to assign a respondent to a single category (whole probabilistic methods), this cumbersome process can be avoided. The All Inclusive method, like the Equal Fractional method, does not depend on the particular distribution, but it does produce proportions that add to more than 100 percent unless they are raked or recalculated to a base of 100 percent each time. The Deterministic Whole Assignment methods and the NHIS Fractional method would require an extra step unless only national figures are used, because the relative size of the groups must be determined for each detailed distribution. Otherwise, they are as easy to use as the whole probabilistic methods.

Skill Required. This criterion refers to the skills required to carry out the bridge operations. The amount of computer expertise to perform the operations associated with each of these methods is fairly trivial. The Deterministic Whole Assignment methods require almost no statistical knowledge. Some familiarity with the statistical adjustment literature would be useful for understanding the Deterministic Fractional Assignment procedures. If the All Inclusive method were used, users might need to understand statistical raking.

Understandability and Communicability. This criterion concerns how easily the methods can be explained and understood by the average user. The Deterministic Whole Assignment methods are both easy to explain and easy to understand. The fractional assignment of individuals to a single category also is not difficult to follow. Assigning fractions of a person to different categories may be easy to explain, but the average user may find it difficult to accept the idea. The All Inclusive method also is easily explained, but, unless the percentages are raked to 100 percent, users may have a problem understanding how to use the results.

E. Weighting When Appropriate Population Controls Are Not Available

For those using the new racial categories in surveys prior to the release of new population controls from Census 2000 (expected in 2002), a method is needed to allow the use of the updated 1990 controls. The following advice is provided for researchers who find themselves in this situation:

1. Choose a whole-allocation bridging method.
2. Create a bridged distribution using the chosen method.
3. Rake this bridged distribution to the 1990-based controls.
4. Use the final weight from this process when reporting distributions for the new racial categories, including the multiple-race combinations.
5. Provide the following caveats to data users: (a) weighting was not done using controls based on the new definitions, and (b) the bridged distribution is not necessarily the same as would have been obtained using the old race question.

Choosing a whole-allocation bridging method will simplify the task. Data producers should select the bridging method that they judge to be the most appropriate for their data users. Provide any available information to evaluate the likely discrepancies between the bridged distribution and the distribution that would have been obtained using the old question. Reweighting to the new controls, once they become available, is strongly recommended.

F. Strategies for Users

This section suggests strategies for bridging that depend on the user's needs. In providing these strategies, two assumptions are made. The first is that the user is interested in the analysis of an historical data series. The second assumption follows from the first, and it is that the user will not simply be doing cross-sectional estimates of the current counts in the racial categories or the present and future characteristics of these populations. It should be noted that all of the bridging methods would require, to some extent, prior knowledge of population characteristics when working at the local level.

If the user wants to examine change across the whole racial distribution, the Plurality method or one of the Fractional Allocation methods will likely provide the best approximations to the distributions from the past. They actually are based on information concerning how the respondents would have answered the question on race under the previous standards. While the new distribution should reflect only the growth in the racial categories due to population change and not methodological change, most of the other methods produce a substantial increase in the size of the American Indian and Alaska Native population not attributable to change over time. Accompanying the increase in the AIAN category is a corresponding decrease in the White category. Other distortions could occur as the size of the population identifying with more than one race grows.

If the user is interested in a particular racial category, which method is chosen will depend on whether the user wants to err on the side of inclusion or exclusion, especially for the smaller racial groups. Use of the Smallest Group method or the Largest Group Other Than White method will include a substantial number of individuals that might have identified with the larger racial group in the past. The characteristics of the group might change simply as a result of their inclusion. This is particularly the case for the AIAN category, and the differences between methods with respect to inclusion or exclusion will be greatest in areas having a large AIAN population, such as the state of Washington.

Analysis of the combined race/ethnicity format indicates that the choice of tabulation method has little effect on Hispanics. It should be remembered, however, that these results are based on the use of a "manufactured" variable and may not truly represent what would occur if an actual combined format had been used in the past. It is also the case that the AIAN category is still affected by the choice of bridging method when using the combined format.

The choice of bridging method also could depend on the substantive characteristics of interest. Some characteristics may not be affected by the particular bridging method chosen. Others will show more change with one method or another. In these cases, which bridging method is used

will depend upon what the user is trying to discover. As noted above, health outcomes appeared not to be affected by the choice of bridging method, but economic outcomes were.

Geographic characteristics also could affect the choice of bridging method. The user may have knowledge of local populations that would dictate which method should be used. For instance, the user might know that in a particular area Whites also are likely to identify as American Indian when given the opportunity, but, if forced to select a single race, they would almost certainly choose White. This knowledge might lead the user to use a bridging method that minimizes assignment to the American Indian and Alaska Native category.

The fact that the Smallest Group method and the Largest Group Other Than White method give larger counts for the smaller racial groups might favor their selection in some circumstances. Reliability and confidentiality standards normally would restrict the analysis of these categories. Even though these bridging methods may produce less than ideal comparisons to the past, the larger size of the categories resulting from using these methods could increase analytical power enough to draw conclusions.

Although the selection of a bridging method should be for substantive or methodological reasons, simplicity cannot be discounted altogether. Users must have both the substantive knowledge and methodological skill to use the more complicated methods. They also must be able to explain them and defend them.

References

Benson, V. and Marano, M. (1995), "Current Estimates from the National Health Interview Survey, 1994," National Center for Health Statistics, *Vital Health Statistics*, 10(193).

Massey, J. T., Moore, T. F., Parsons, V. L., and Tadros W. (1989), "Design and Estimation for the National Health Interview Survey, 1985-1994," National Center for Health Statistics, *Vital Health Statistics*, 2(110).

Rubin, D. R. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley, 1987.

Tucker, C., McKay, R., Kojetin, B., Harrison, R., de la Puente, M., Stinson, L., and Robison, E. (1996), "Testing Methods of Collecting Racial and Ethnic Information: Results of the Current Population Survey Supplement on Race and Ethnicity," *Bureau of Labor Statistics Statistical Notes*, No. 40.