

The State of Federal Health Statistics on Racial and Ethnic Groups

Robert A. Hahn, PhD, MPH

Objective.—To examine assumptions underlying federal health statistics on racial and ethnic groups in the United States.

Data Sources.—Studies conducted by federal agencies and other investigators, and technical appendices of published vital statistics and census reports.

Data Synthesis.—Several assumptions underlying federal health statistics on racial and ethnic groups are not well supported. Conceptual (as opposed to operational) definitions of race and ethnicity are not available, and scientific grounds for definition are not considered. Procedures for the ascertainment of race and ethnicity vary within and among data-collection agencies. Miscounting and misclassification may vary by an order of magnitude between whites and other races. The responses of individuals to questions of racial and ethnic identity differ for different indicators, in different surveys, and at different times. As a result, counts, rates, and rate ratios may not be meaningful or accurate. Particularly for Hispanics and for races other than whites or blacks, there are inconsistencies in statistical information that may hinder health research and program development.

Conclusions.—Improvement of federal health statistics for racial and ethnic groups requires (1) clarification of goals for classification, (2) adoption of scientific principles for the validation and definition of the categories "race" and "ethnicity," (3) assessment of perceived social identity in the population, and (4) periodic evaluation.

(JAMA: 1992;267:268-271)

FEDERAL health statistics indicate that infant mortality in the United States in 1987 was 2.1 times higher among blacks and 1.2 times higher among American Indians than among whites,¹ that cumulative rates of the acquired immunodeficiency syndrome through May 1988 were 2.9 times higher among non-Hispanic blacks and 2.6 times higher among Hispanics than among non-Hispanic whites,² and that the proportion of persons not covered by any health insurance in 1986 was 1.6 times greater among blacks than among whites.³

See also pp 259 and 275.

The validity of such health statistics rests on logical assumptions that include the following: (1) The categories of "race" and "ethnicity" and specific racial and ethnic group designations are consistently defined and ascertained. (2) The categories and designations are understood by the populations questioned. (3) Survey enumeration, participation, and response rates are high and similar for all populations. (4) The responses of individuals are consistent in different data sources and at different times. This report builds on previous research and examines these assumptions, focusing on birth, death, and population statistics in the United States.

Health statistics on racial and ethnic groups in the United States are calculated from data on the population provided by the Bureau of the Census (BC) and data on health events provided by the Centers for Disease Control's National Center for Health Statistics (NCHS). The BC and NCHS have conducted many studies to assess and enhance the quality of their data. This report uses results of BC, NCHS, and other important studies conducted since the 1950s, recognizing that the system of data collection has evolved substantially over this period.

In the collection of information on race and ethnic groups, consistency of definition and procedure is not simply a matter of convenience for analysts. Sources of information on natality, mortality, and the population are highly interdependent. Birth and death rates are calculated on the basis of the census,^{1,4} completeness of birth registration is assessed by census information,⁵ intercensal estimates and estimates of completeness of coverage in the decennial census require information on births and deaths,⁶ and infant mortality is computed as a ratio of infant deaths to infant births.¹

Long-standing conceptual difficulties in the definition of "race" and "ethnicity" pose a challenge to the surveillance of health in US "racial" and "ethnic" groups. The validity of the concept of "race," for example, has been questioned,^{7,8} and biological notions of "race" may be confused with cultural and behavioral notions of "ethnicity." Lack of scientific consensus has hindered the establishment of firm principles for surveillance of the health of racial and ethnic groups. Surveillance is also made difficult by multiple and changing perceptions of social identity in different segments of the population and by interests not consistent with scientific goals. While data-collection agencies have made notable improvements, this review suggests a need both to reconsider the scientific validity of the categories, "race" and "ethnicity," and to assess perceptions of social identity in the US population.

FOUR ASSUMPTIONS EXAMINED

Assumption 1: The Categories of 'Race' and 'Ethnicity' Are Consistently Defined and Ascertained by Federal Data-Collection Agencies

Background.—Current regulations for the statistical classification of racial and ethnic groups by federal agencies were published in 1978 in Office of Management and Budget (OMB) directive

From the Epidemiology Program Office, Centers for Disease Control, Atlanta, Ga.

Reprint requests to Epidemiology Program Office, C-08, Centers for Disease Control, Atlanta, GA 30333 (Dr Hahn).

15, "Race and Ethnic Standards for Federal Statistics and Administrative Reporting."¹⁰ Directive 15 was intended to set uniform standards for data collection and publication among federal agencies and to increase available information on persons of Hispanic origin, as required by Congress.^{11,12} Directive 15 does not define the concepts of "race" or "ethnicity" and explicitly denies being "scientific or anthropological in nature." It presents brief rules for the classification of persons into four racial categories (American Indian or Alaskan Native, Asian or Pacific Islander, black, and white) and two ethnic categories (Hispanic origin and not of Hispanic origin); it requires categorization of whites and blacks as Hispanic or non-Hispanic.

Directive 15 employs four features to define categories of race and ethnicity: (1) descent from "the original peoples" of a specified region or nation, (2) a specific cultural origin, (3) cultural identification or affiliation, and (4) race itself. Defining characteristics are not further explained. Descent is used to define all six categories (of race and ethnicity together). Specific cultural origin is used to define Hispanics; however, American Indians and Alaskan Natives are similarly defined by "cultural identification through tribal affiliation or community recognition." "Black race" is tautologically defined by black race and African origin. The race and ethnicity categories of directive 15 are not mutually exclusive; for example, "Hispanic" is not exclusive of either "American Indian or Alaskan Native" or of "Asian or Pacific Islander." Supporting documents indicate that the categories of directive 15 were not intended to be exhaustive.¹¹

Directive 15 does not clearly indicate whether race and ethnicity are to be ascertained on the basis of the respondent's or the interviewer/observer's perception. While the directive suggests that the respondent's self-perception or perhaps "recognition in his community" determines racial and ethnic identity, supporting documents imply that self-identification is not intended by the directive and is not the optimal method of ascertainment.¹¹ The difference between respondent and interviewer/observer ascertainment may be substantial, both conceptually and statistically. An NCHS study that compared respondent-reported race with interviewer-observed race in a nationally representative sample found that 5.8% of persons who reported themselves as "black" were classified as "white" by the interviewer and that 32.3% of self-reported Asians and 70% of self-reported American Indians were classified as "white" or "black."¹³

Definitions of 'Race' and 'Ethnicity' in NCHS and BC Statistics.—As in directive 15, NCHS and BC documents do not define the concepts of "race" and "ethnicity" or consider the scientific principles on which classification schemes are based. ("The concept of race as used by the Census Bureau reflects self-identification; it does not denote any clear-cut scientific definition of biological stock."¹⁴) In addition, the terminology and categorization of racial and ethnic groups differ from source to source. For example, the category prescribed by directive 15, "American Indians," is labeled "Indian (Amer)" in the 1980 census and lists separately Eskimos and Aleuts; in natality documents, the category "American Indian" includes Aleuts and Eskimos; and in mortality documents, the category "Indian" includes American, Canadian, Eskimo, and Aleut.^{14,14} Categorization of "American Indians" (and of other groups) is thus not strictly commensurate among sources and does not fully comply with directive 15.

Birth and Death Statistics.—Procedures used to assign race and ethnicity also differ among sources of information. Until 1989, the race of a newborn was determined by a complex algorithm incorporating information on the race of the parents.⁴ Races were treated unequally; for example, only infants with two white parents were white, while an infant with one white parent took the race of the parent whose race was other than white.⁴ Beginning in 1989, while information on the race of the parents will still be available in computer files, published NCHS statistics uniformly assign the infant the race and ethnicity of its mother, diminishing the number of infants assigned to races other than white while increasing the number of infants assigned to the white race.¹⁵ Statistics on Hispanic origins, published since 1984, have always assigned newborns the origin of their mothers.

On death certificates, parental race and ethnicity are not cited as definitional criteria, other than for fetal deaths. Race and ethnicity are determined by information about the decedent's identity given to funeral directors by next of kin¹; however, with unknown frequency, funeral directors make independent assessments.

The Census.—In the decennial census, respondents are asked to indicate their own racial and ethnic identity and that of household members.¹⁴ Postcensal and intercensal population estimates incorporate information from several sources: previous and subsequent decennial censuses, subsequent registered

births and deaths, and migration (ascertained from the Immigration and Naturalization Service by an algorithm including demographic information on the immigrants' nations of origin). Intercensal estimation procedures tacitly (and erroneously) assume that race and ethnicity are commensurate in all sources.⁶ Evidence presented below indicates that individuals may be assigned a different race at birth, during the course of life, and at death.

Assumption 2: Racial and Ethnic Categories Are Understood by the Populations Questioned

While popular understandings of race and ethnicity have not been comprehensively explored, there are indications that popular notions differ substantially from those of information-collection agencies. Interviewers in an NCHS survey, for example, report that "the phrase 'origin or descent' was poorly understood by many respondents,"¹⁶ and the BC remarks that notions of "race," "ethnicity," and "ancestry" are not clearly distinguished from one another by census respondents.¹⁷ The BC assumes it has asked meaningful questions when response rates are high¹⁷; however, survey respondents may answer questions they find unclear.¹⁸ The BC also assumes its results are valid when overall counts are similar in reinterview surveys,¹⁹ but overall counts may be similar even when large proportions of respondents have changed their responses.

Terminology for race and ethnic groups also differs among segments of the population. The category "white" is understood by some Hispanics to be synonymous with "Anglos," ie, non-Hispanics. Perhaps for this reason, in the 1980 census, almost 40% of persons who classified themselves as Hispanic answered "other" to the question intended to elicit racial identity.²⁰ (The 1980 census question intended to elicit "race" did not use the word *race*, causing confusion for both respondents and analysts.) While 1980 census documents report these persons as "other," intercensal estimates have reassigned them a race in proportion to Hispanics who specified "black" or "white" race, to comply with requirements of OMB directive 15.²⁰ However, the BC also estimates that, among persons (in 27 states) who specified a race, 32.7% of whites and 92.9% of blacks who reported being Mexican-American did so erroneously; thus, approximately 33.4% of the estimated Mexican-American population in the states examined was misclassified.²¹ Thus, persons who apparently did not find an appropriate option on the "race" question have been assigned a race on the basis of persons

who answered the question, but commonly in error.

Other questionnaire terms are also misinterpreted. The category "South and Central American" (in the 1970 census) was thought by some respondents to refer to natives of southern and central regions of the United States, and "Alaska Native" was thought to refer to persons born in Alaska.¹⁷ Finally, categories such as "Other" and "Mexican-(Amer)" that appeared in the 1980 census are also reported to be misunderstood, particularly by recent immigrants and minority respondents.¹⁷

Assumption 3: Survey Enumeration, Participation, and Response Rates Are High and Similar for All Racial and Ethnic Populations

Underregistration of Births.—Underregistration of births overall appears to be negligible but varies by race. The most recent survey estimated that birth underregistration between 1964 and 1968 was 0.6% for whites and 2.0% for "Negro and other races."¹⁶ However, the inclusion of blacks and other groups in a single category may mask diversity. The earlier survey of births in 1950 found an underregistration of 1.4% for whites, 6.3% for blacks, and 14.9% for American Indians, indicating a marked improvement from 1950 to 1964-1968 but also indicating a marked racial disparity.²²

Misclassification of Race at Death.—In the registration of deaths, the problem is less one of underregistration than of misclassification—incorrectly assigning race to decedents. The net result is similar to underregistration and undercounting—deaths are undercounted in one race and overcounted in another. Study of a sample of 1986 US death certificates indicated that, while white race was misclassified on only 0.5% and black race on only 1.0% of certificates, Asians were misclassified on 21.1% of certificates and American Indians on 23.7%.²³ Most misclassified decedents were falsely reported to be white, exaggerating white mortality and minimizing mortality among races other than white. Misclassification appears not to have decreased in the last 25 years; similar results have been found in studies of US deaths in 1960²⁴ and infant deaths in Washington State between 1968 and 1977.²⁵

Census Enumeration Miscounts.—Differential miscounting by race and ethnicity has also occurred in the census. It is estimated that the 1980 census undercounted blacks by 5.9% and whites and others by 0.7%; Hispanics were undercounted by as much as 7.8%.⁶ The

1980 census has been shown to have "overcounted" American Indians by approximately 358 000, 33.7% above the population estimated from previous censuses and intervening births and deaths; the overcount is probably a result of increasing "Indian" self-identification.²⁶

Nonresponse.—A small proportion of nonresponses to racial identity questions on birth and death certificates is imputed by an algorithm that assigns the decedent the race of the preceding computer record, while nonresponse on Hispanic origins on infant death certificates is reported to be as high as 8%.¹ In 1988, the most recent year of published mortality statistics, only 18 states (with an estimated 80% of the Hispanic population) met NCHS standards for the reporting of infant mortality by Hispanic origin.^{1,4}

Nonresponse to the "race" question on the 1980 census was only 1.5%; however, as noted, 40% of persons identifying themselves as Hispanic (almost 6 million persons) chose the "other" category on this question. The Hispanic-origin question on the 1980 census was unanswered by 2.3% of respondents, the ancestry question by 12%.^{17,27}

The effects of miscounts and misclassification may be substantial.⁶ For example, computer-linked statistics on US infants dying in 1983 and 1984 indicate that infant mortality was 1.57 times higher among American Indians than among whites (using the uniform rule assigning a newborn and decedent infant the race of its mother).²⁸ This is almost 50% higher than the 1.06 ratio based on published information in which infants may be assigned a different race at birth and at death.^{29,30}

Assumption 4: Individual Responses to Questions of Racial and Ethnic Identity Are Consistent in Different Surveys and Different Times

Given that persons of different racial and ethnic groups are covered by surveys and have responded to questions about "race" and "ethnicity," available evidence indicates substantial inconsistency in responses. The BC researchers Siegel and Passel³¹ suggest that inconsistency may be explained by two phenomena: "fuzzy group boundaries," ambiguity about the criteria of group membership, and "shifting identity," changes of individuals' group identity over time.

Siegel and Passel's study of coverage in the 1970 census compared five indicators of self-reported Hispanic identity: Hispanic origins, use of the Spanish language, Hispanic heritage, surname, and birthplace or parentage.³¹ With different indicators, estimates of the Hispanic population ranged from

5.2 million (persons of Hispanic heritage) to 1.8 times as many, 9.6 million (persons using the Spanish language). The BC researchers conclude, "A central problem is the inability of the census data to reflect a clear, unambiguous, and objective definition of exactly who is a member of the Hispanic population."

The category "American Indian" is also characterized by boundary fuzziness and shifting identification, as indicated above by apparent population "overcounts." In a 1980 reinterview study, 41% of persons who identified themselves as American Indian had reported themselves as white in the 1980 census.¹⁹ Also in the 1980 census, while an estimated 6.8 million persons reported American Indian ancestry, only 1.2 million reported American Indian race. Ancestry was not simply a broader category; only 73% of those reporting American Indian race claimed American Indian ancestry, the remainder claiming European, other, or unknown ancestries.³² In questioning about perceived identity, the circumstances and terminology used appear to have important effects on responses.

Shifting perception of identity has also been found in other race and ethnic groups. In answer to the 1980 census "race" question, 26.5 million people identified themselves as "black or Negro," while only an estimated 21 million claimed Afro-American ancestry.³³ In March 1971 and again in March 1972, the BC's Current Population Survey interviewed a large sample of persons in US households, eliciting the ethnic identity of all household members; from one year to the next, 34.3% of household members were reported to have different ethnic identities.³⁴

COMMENT

Because of their frequent association with health status, "race" and "ethnicity," along with age and sex, are categories by which populations are commonly divided in health research and planning. Such categorization assumes that "race" and "ethnicity" are valid concepts that can be correctly identified and classified. The validity of the concept of "race" has been questioned in recent decades.⁷⁻⁹ Even if conceptually valid, it may be that race cannot be readily assessed by survey procedures.

Particularly for nonwhite populations and for Hispanics, the quality of statistical information on the health of racial and ethnic populations in the United States is problematic. Better-quality information would enable more effective health research and program planning. While the system of health statistics on racial and ethnic groups has not been

comprehensively evaluated, available evidence suggests that certain underlying assumptions do not always hold. Conceptual definitions are lacking and the potential scientific bases for definition are largely ignored; procedures for data ascertainment vary from source to source; undercounting and misclassification can differ by orders of magnitude between whites and races other than white or black; and the responses of individuals may be inconsistent for different indicators, in different surveys, and at different times. In information systems that rely on separate sources, inconsistent definitions and procedures for assigning race and ethnicity may produce computed results of questionable validity. Counts, rates, and rate ratios may be inaccurate.

Further improvement of the system of health statistics may require several steps:

1. Clear goals (and possibly diverse ones) for the definition of categories of "race" and "ethnicity" need to be established for planning, public health surveillance, administration, and research.

2. Extensive efforts should be made to scientifically validate the categories

"race" and "ethnicity" and to establish basic scientific and anthropological principles for the public health surveillance of racial and ethnic populations. In the validation of "race," genetic, physical anthropological, archaeological, linguistic, and migration evidence should be weighed. In the validation of "ethnicity," concepts of social identity and changes in social identity in various populations should be assessed and compared. If valid, categories must be clearly defined, standardized, and operationalized so that researchers and respondents know what the categories mean and so that information is compatible from source to source. The design, development, and management of programs and administrative activities should also be based on scientific principles.

3. To ensure that respondents to public health surveillance understand survey instruments, the way in which different segments of the population identify themselves by "race," "ethnicity," or other characteristics should be comprehensively assessed and incorporated into surveillance design.

4. To ensure the quality of health sta-

tistics of racial and ethnic populations, the surveillance system should be periodically and systematically evaluated.³⁵

The assessment of demographic identity in a society of culturally diverse and rapidly changing populations with different needs and interests is extremely difficult. In meeting this challenge, federal agencies have made notable improvements, for example, the new NCHS rule assigning infants the race of their mother in published statistics and the NCHS linked birth/infant-death computer tape. A Public Health Service Task Force on Minority Health Data has recently been constituted to further address data problems and needs. With additional collaborative work by federal, state, and local organizations in the collection, analysis, and reporting of population and health statistics, substantial progress toward resolution of the problems outlined in this report is possible.

I am grateful for the useful suggestions and comments of J. Gregory Robinson, PhD, and Arthur R. Cresce, PhD, of the Bureau of the Census, Ronald W. Wilson, MA, of the National Center for Health Statistics, Hyattsville, Md, and Steven M. Teutsch, MD, and Joseph Mulinare, MD, of the Epidemiology Program Office, Centers for Disease Control, Atlanta, Ga.

References

1. *Vital Statistics of the United States 1987: Mortality*. Washington, DC: US Dept of Health and Human Services; 1990;IIA. Publication PHS 90-1101.
2. Selik RM, Castro KG, Pappaianou M. Racial/ethnic differences in the risk of AIDS in the United States. *Am J Public Health*. 1988;78:1539-1545.
3. *Health, United States, 1989*. Hyattsville, Md: National Center for Health Statistics; 1990.
4. *Vital Statistics of the United States 1987: Natality*. Washington, DC: US Dept of Health and Human Services; 1989;I. Publication PHS 89-1100.
5. *1970 Census of Population and Housing: Evaluation and Research Program: Test of Birth Registration Completeness 1964 to 1968*. Washington, DC: Bureau of the Census; 1973. US Dept of Commerce publication PHC80-E-2.
6. Fay RE, Passel JS, Robinson GJ, Cowan CD. *The Coverage of the Population in the 1980 Census*. Washington, DC: Bureau of the Census; 1988. US Dept of Commerce publication PHC80-E-4.
7. Montagu A. *The Idea of Race*. Lincoln, Neb: University of Nebraska Press; 1965.
8. Lewontin RC. The apportionment of human diversity. *Evol Biol*. 1972;6:381-398.
9. Gould SJ. Why we should not name human races: a biological view. In: Gould SJ, ed. *Ever Since Darwin*. New York, NY: WW Norton & Co Inc; 1977:231-236.
10. Office of Management and Budget. Directive No. 15: race and ethnic standards for federal statistics and administrative reporting. In: *Statistical Policy Handbook*. Washington, DC: Office of Federal Statistical Policy and Standards, US Dept of Commerce; 1978:37-38.
11. Wallman KK, Hodgdon J. Race and ethnic standards for federal statistics and administrative reporting. *Stat Rep*. 1977;77:450-454.
12. Wallman KK. Statistics for Americans of Spanish origin or descent. *Stat Rep*. 1978;78:148-152.
13. Massey J. Using interviewer observed race and respondent reported race in the Health Interview Survey. In: *Proceedings of American Statistical Association Meetings: Social Statistics Section*. Alexandria, Va: American Statistical Association; 1980:425-428.
14. *1980 Census of Population: General Social and Economic Characteristics: United States Summary*. Washington, DC: Bureau of the Census; 1988. US Dept of Commerce publication PC80-1-C.
15. Heuser RL. Race and ethnicity in U.S. natality tabulations. Presented at the 117th Annual Meeting of the American Public Health Association; October 24, 1989; Chicago, Ill.
16. Drury TF, Moy CS, Poe GS. Going beyond interviewer observations of race in the National Health Interview Survey. In: *Classification Issues in Measuring the Health Status of Minorities*. Hyattsville, Md: National Center for Health Statistics; 1980:5-17.
17. McKenney NR, Cresce AR, Johnson PA. U.S. development of the race and ethnic items for the 1990 census. Presented at the Annual Meeting of the Population Association of America; April 21-23, 1988; New Orleans, La.
18. Suchman L, Jordan B. Interactional troubles in face-to-face survey interviews. *J Am Stat Assoc*. 1990;85:232-244.
19. *Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview*. Washington, DC: Bureau of the Census; 1986. US Dept of Commerce publication PHC80-E-2.
20. Hollmann FW. *United States Population Estimates, by Age, Sex, Race, and Hispanic Origin: 1980 to 1988*. Washington, DC: Bureau of the Census; 1990. US Dept of Commerce, ser P-25, No. 1045.
21. *Persons of Spanish Origin by State: 1980*. Washington, DC: Bureau of the Census; 1982. US Dept of Commerce publication PC80-S1-7.
22. *Vital Statistics of the United States 1950*. Washington, DC: US Dept of Health, Education, and Welfare; 1954;I.
23. Poe GS, Powell-Griner E, McLaughlin JK, Robinson K, Placek J. A comparability study of the responses to the 1986 National Mortality Followback Survey to similar demographic items from the death certificate. Unpublished results.
24. Hambricht TZ. Comparability of marital status, race, nativity, and country of origin on the death certificate and matching census record: United States—May-August 1960. *Vital Health Stat [2]*. 1969;1-47. Ser 2, No. 84.
25. Frost F, Shy KK. Racial differences between linked birth and infant death records in Washington State. *Am J Public Health*. 1980;70:974-976.
26. Passel JS, Berman PA. Quality of the 1980 census data for American Indians. *Soc Biol*. 1986;33:163-182.
27. Farley R. *Race and Ethnicity in the U.S. Census: An Evaluation of the 1980 Ancestry Question*. Washington, DC: Bureau of the Census; 1990.
28. Kleinman, JC. Infant mortality among racial/ethnic minority groups, 1983-1984. *MMWR*. 1990;39(SS-2):31-39.
29. *Vital Statistics of the United States 1983: Mortality*. Washington, DC: US Dept of Health and Human Services; 1987;IIA. Publication PHS 87-1101.
30. *Vital Statistics of the United States 1984: Mortality*. Washington, DC: US Dept of Health and Human Services; 1987;IIA. Publication PHS 87-1122.
31. Siegel JS, Passel JS. Coverage of the Hispanic population of the United States in the 1970 census. Washington, DC: Bureau of the Census; 1979. Current Population Reports. US Dept of Commerce publication P23, No. 82.
32. Snipp CM. Who are American Indians? some observations about the perils and pitfalls of data for race and ethnicity. *Popul Res Policy Rev*. 1986;5:237-252.
33. *1980 Census of Population: Ancestry of the Population by State: 1980*. Washington, DC: Bureau of the Census; 1983. Current Population Reports. US Dept of Commerce publication PC80-S1-10.
34. Johnson CE Jr. *Consistency of Reporting of Ethnic Origin in the Current Population Survey*. Washington, DC: Bureau of the Census; 1974. US Dept of Commerce Technical Paper No. 31.
35. Centers for Disease Control. Guidelines for evaluating surveillance systems. *MMWR*. 1988;37(S-5):1-18.