

Ethnicity & Health, 1999; 4(3): 121-137



The Monitoring of Racial/Ethnic Status in the USA: Data Quality Issues

DAVID R. WILLIAMS PhD, MPH

University of Michigan, Ann Arbor, USA

ABSTRACT

This paper reviews the assessment of racial and ethnic identification in the major data collection systems of the US Department of Health and Human Services. It evaluates the quality of the available data and outlines recommendations for improving the collection of racial data and enhancing our understanding of the role of race in health. Special attention is also given to the role of socioeconomic status in understanding racial differences in health and the assessment of racial status in data systems in the UK.

Keywords: Ethnicity, socioeconomic status, race.

INTRODUCTION

Race has not been measured in a consistent manner over time in the USA. Race has been measured in the decennial census since its inception but the Census Bureau has routinely changed its racial categories over the last two centuries, with no racial classification scheme having been used in more than two censuses.¹ The Federal government's Office of Management and Budget (OMB) current guidelines for measuring race and ethnicity in the USA recognizes four racial groups (white, black, Asian or Pacific Islander [API], and American Indian or Alaskan Native) and one ethnic category (Hispanic) in the USA. This directive has been in effect since 1977. Recently, the OMB has proposed new guidelines that must be implemented in all federal statistical programs by the year 2003. These new guidelines recognize five racial groups (white, black or African American, Asian, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native) and one ethnic category (Hispanic or Latino). In addition, the new guidelines will allow persons of mixed racial ancestry to list themselves in as many racial categories as apply. Standardized tabulation procedures for counting persons who identify with more than one race have not yet been developed.

From the very beginning, racial categories in the USA reflected a hierarchy of racial preference that was driven by a racist ideology. Three of the four currently recognized racial categories were in the very first census in 1790 and they were not regarded as equal. In compliance with Article One of the US Constitution, this census enumerated whites, blacks as three-fifths of a person and only those Indians who paid taxes. The Thirteenth Amendment abandoned the Three-Fifths Rule, but Indians continued to be divided into the categories of 'civilized Indians' and 'Indians not taxed' until 1924 when all American Indians were granted US citizenship by Congress.²

Correspondence to: Dr D. R. Williams, University of Michigan, Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, MI 48106-1248, USA.

TABLE 1. *Racial/ethnic measurement in selected surveys and major data collections, US Department of Health and Human Services*

Name	Data source/method	Uses of data	Sample	Race/ethnicity and socioeconomic status	Periodicity
I. Population-based surveys					
National Survey of Family Growth (NSFG)	<ul style="list-style-type: none"> Personal interview 	<ul style="list-style-type: none"> Contraception and sterilization Teenage sexual activity and pregnancy Family planning and unintended pregnancy Adoption Breastfeeding Infertility 	<ul style="list-style-type: none"> 14 000 women 15-44 years of age sampled to complete over 10 000 interviews Oversample blacks and Hispanics 	<ul style="list-style-type: none"> OMB categories Four Hispanic groups Family and individual income and poverty level Sources of income Education and occupation 	3 and 4 years (2000 and 2003)
National Health Interview Survey (NHIS)	<ul style="list-style-type: none"> Personal interview 	<ul style="list-style-type: none"> Annual data on: <ul style="list-style-type: none"> Health status and disability Utilization of health care AIDS knowledge and attitudes Family resources Health insurance Access to care Immunization Injury Health behaviors Functioning 	<ul style="list-style-type: none"> 41 000 households Oversample blacks and Hispanics 	<ul style="list-style-type: none"> OMB categories Four Hispanic groups Family and individual income and poverty level Education and occupation Type of living quarters 10 API groups 	Annual
II. Record-based surveys					
National Hospital Discharge Survey (NHDS)	<ul style="list-style-type: none"> Hospital records Computerized data sources 	<ul style="list-style-type: none"> Patient characteristics Length of stay Diagnosis and multiple diagnoses Surgical and diagnostic procedures 	<ul style="list-style-type: none"> 542 hospitals 250 000 discharges 	<ul style="list-style-type: none"> OMB categories 	Annual
National Survey of Ambulatory Surgery (NSAS)	<ul style="list-style-type: none"> Abstract forms completed by facility staff 	<ul style="list-style-type: none"> Patient characteristics Diagnosis and multiple diagnoses Surgical and diagnostic procedures 	<ul style="list-style-type: none"> 750 facilities 180 000 patients 	<ul style="list-style-type: none"> OMB categories 	Annual through 1996; periodic thereafter
National Ambulatory Medical Care Survey (NAMCS)	<ul style="list-style-type: none"> Encounter forms completed by physicians practicing in private offices 	<ul style="list-style-type: none"> Characteristics of patients' visits to physicians Diagnoses and treatment 	<ul style="list-style-type: none"> 3000 physicians in office-based practices 45 000 patient visits 	<ul style="list-style-type: none"> OMB categories 	Annual

National Hospital Ambulatory Medical Care Survey	<ul style="list-style-type: none"> • Encounter forms completed by physicals and other hospital staff 	<ul style="list-style-type: none"> • Characteristics of patients' visits to hospital outpatient departments and emergency departments • Diagnoses and treatment • Number and characteristics of residents in nursing homes (NH) • Functional status of NH residents • People discharged from NH • Expenditures of NH • Number of patients • Functional status of patients • Number of discharged patients • Characteristics of home health agencies and hospices and their patients 	<ul style="list-style-type: none"> • 600 hospitals • 90 000 patient visits 	<ul style="list-style-type: none"> • OMB categories 	Annual
National Nursing Home Survey (NNHS)	<ul style="list-style-type: none"> • Long-term care providers • Resident next-of-kin interviews 	<ul style="list-style-type: none"> • Number and characteristics of residents in nursing homes (NH) • Functional status of NH residents • People discharged from NH • Expenditures of NH • Number of patients • Functional status of patients • Number of discharged patients • Characteristics of home health agencies and hospices and their patients 	<ul style="list-style-type: none"> • 1200 nursing homes (NH) • 7200 NH residents • 7200 NH discharges 	<ul style="list-style-type: none"> • OMB categories 	Historically every 4-10 years; converting to bi-annual cycle
National Home and Hospice Care Survey (NHHCS)	<ul style="list-style-type: none"> • Home health agencies and hospices 	<ul style="list-style-type: none"> • Number of patients • Functional status of patients • Number of discharged patients • Characteristics of home health agencies and hospices and their patients 	<ul style="list-style-type: none"> • 1200 home health agencies and hospices • 7200 current patients • 7200 discharged patients 	<ul style="list-style-type: none"> • OMB categories 	Annual through 1996, bi-annual thereafter
III. Vital statistics system					
Vital Statistics Cooperative Program (VSCP)	<ul style="list-style-type: none"> • State vital registration 	<ul style="list-style-type: none"> • Life expectancy • Causes of death • Infant mortality • Perinatal care and birth weight • Birth rates • Nonmarital births • Pregnancy outcomes • Occupational mortality • Teenage pregnancy 	<ul style="list-style-type: none"> • All births, deaths and fetal deaths • Counts of marriages and divorces 	<ul style="list-style-type: none"> • For births, deaths and fetal deaths: <ul style="list-style-type: none"> • White, black, five API groups, American Indian, Other • Five Hispanic groups • Education • Births and deaths 10 API groups from seven states and NYC • Deaths only: <ul style="list-style-type: none"> • Occupation in 21 states • Same race groups at VSCP • Education • Same race groups as VSCP 	Annual
Linked Birth/Infant Death Program	<ul style="list-style-type: none"> • Birth and Death Certificates 	<ul style="list-style-type: none"> • IM rates by birth and period cohorts • IM rates by birth weight • Facilitates epidemiological follow-up studies • Verification of death for individuals under study • Most NCHS surveys are linked to the NDI 	<ul style="list-style-type: none"> • All US births and infant deaths • All deaths 	<ul style="list-style-type: none"> • Occupation in 21 states • Same race groups at VSCP • Education • Same race groups as VSCP 	Annual
National Death Index (NDI)	<ul style="list-style-type: none"> • State registration death certificates 	<ul style="list-style-type: none"> • IM rates by birth and period cohorts • IM rates by birth weight • Facilitates epidemiological follow-up studies • Verification of death for individuals under study • Most NCHS surveys are linked to the NDI 	<ul style="list-style-type: none"> • All US births and infant deaths • All deaths 	<ul style="list-style-type: none"> • Occupation in 21 states • Same race groups at VSCP • Education • Same race groups as VSCP 	Annual

API, Asian or Pacific Islander; OMB, Office of Management and Budget.

Race has been a fundamental organizing principle of US society.³ Historically, attitudes and beliefs about racial groups have been translated into policies and societal arrangements that limited the opportunities and life chances of stigmatized groups. Minority populations' disproportionate representation at the lower levels of socioeconomic status (SES) reflects the successful implementation of social processes that were designed to relegate groups with undesirable physical characteristics such as skin color to positions and roles consistent with the dominant society's evaluation of them. Not surprisingly, race has emerged as an important determinant of variations in health.

This paper provides a brief overview of the measurement of racial/ethnic status in selected surveys and major data collections of the US Department of Health and Human Services. It will especially attend to current measurement and methodological issues in the study of racial differences in health in the USA and provide an overview of the current problems with racial data and the critical role of SES for understanding racial/ethnic disparities.

RACIAL/ETHNIC MEASUREMENT IN THE US DATA SYSTEMS

This section will provide a brief overview, with representative examples, of major population-based surveys, record-based surveys and the vital statistics system in the USA. Table 1 provides basic descriptive information for each data source and the availability of racial/ethnic and SES data is noted.

Population-based surveys

National Health Interview Survey. The National Health Interview Survey (NHIS) is a principal source of information on the health of the civilian noninstitutionalized population. The survey, conducted annually since 1957, collects information from approximately 50 000 households and 110 000 people on health status, access to care and insurance, health services utilization, health behaviors and other topics. The survey consists of a set of core data items that are repeated each year and a set of supplements that can change each year to address current health topics.

Core data are collected on the incidence of acute conditions, episodes of persons injured, prevalence of chronic conditions and impairments, restriction in activity due to impairment or health problems, respondent-assessed health status, utilization of health care services involving physician care and short-stay hospitalization, and demographic and socioeconomic characteristics. Data on the four OMB racial categories and Hispanic origin are collected. In addition, sub-group identifiers are collected for 10 API populations and four population groups within the Hispanic category.

National Survey of Family Growth. For more than 20 years, the National Center for Health Statistics (NCHS) has been conducting household interviews of women 15–44 years of age to monitor changes in childbearing practices and to measure reproductive health. The 1973 and 1976 surveys included only never-married women, and the 1982 and 1988 surveys represented all women, 15–44 years of age. In 1990, the 1988 participants were reinterviewed by telephone.

The survey provides data on contraception, infertility, use of family planning and infertility services, sexual activity, family formation, family size and related aspects of maternal and child health such as adoption. Racial/ethnic identifiers are included for the five OMB categories in this survey.

Record-Based Surveys

National Health Care Survey. To meet the health statistics needs of the future, a number of the Center's record-based surveys were merged and expanded into one integrated survey of health care providers called the National Health Care Survey (NHCS). In addition to data on traditional health care settings such as hospitals, the NHCS now provides data on alternative health care settings such as ambulatory surgical centers, hospital outpatient departments, emergency rooms, hospices, and home health agencies. Thus the survey is a source of a wide range of data on the health care field and a significant resource for monitoring health care use, the impact of medical technology, and the quality of care provided to a changing American population.

NHCS was built upon the following current NCHS surveys: the *National Hospital Discharge Survey*, the *National Ambulatory Medical Care Survey*, and the *National Nursing Home Survey*. These were complimented by three new surveys: the *National Survey of Ambulatory Surgery*, the *National Hospital Ambulatory Medical Care Survey*, and the *National Home and Hospice Care Survey*. Data are included for the five OMB categories in each of these surveys.

National Hospital Discharge Survey (NHDS). The National Hospital Discharge Survey is the principal source of information on inpatient utilization of hospitals. Conducted annually since 1965, the survey provides data on the use of non-Federal short-stay hospitals, on their size, location and ownership as well as data on diagnoses, surgical procedures, length of stay, expected source of payment and patient characteristics. Data from the NHDS are useful for tracking specific diseases, the introduction of new technologies, and the impact of changes in financing systems. The survey is currently based on data abstracted from 274 000 records from 525 hospitals.

National Ambulatory Medical Care Survey. The National Ambulatory Medical Care Survey provides data on visits to physicians, including information on patient characteristics, diagnostic procedures, patient management, and planned future treatment. The survey was conducted annually from 1974 to 1981, in 1985, and annually since 1989. Data are collected using encounter forms from approximately 3000 physicians from a sample of 40 000 visits. Data collection from the physician, rather than from the patient, provides an analytic base that expands information on ambulatory care collected through other NCHS surveys.

National Nursing Home Survey. The National Nursing Home Survey provides information on nursing homes from two perspectives—that of the provider of services and that of the recipient. Data about the facilities include characteristics such as size, ownership, Medicare and/or Medicaid certification, occupancy rate, days of care provided, and expenses. For recipients, data are obtained on demographic characteristics, health status, and services received. Conducted periodically since 1963 and most recently in 1995, the survey is based on self-administered questionnaires and interviews with administrators and staff in a sample of about 1500 facilities.

National Survey of Ambulatory Surgery. Although data have been available for three decades for surgery on inpatients, advances in medical technology permit a wide variety of surgical and diagnostic treatments outside the inpatient setting. The National Survey of Ambulatory Surgery provides detailed data on the use of free-standing and hospital-based ambulatory surgery centers in the USA. Data collection began in 1994 and data are available on patient characteristics including age and sex; administrative information including patient disposition, expected sources of payment, and region of the country

where surgery was performed; and medical information including diagnoses, surgical operations, and diagnostic procedures performed.

National Hospital Ambulatory Medical Care Survey. The National Hospital Ambulatory Medical Care Survey produces statistics representing the experience of the US population in hospital emergency departments and outpatient departments. Specifically, the survey provides information on the demographic characteristics of patients, expected source of payment, patients' complaints, physicians' diagnoses, diagnostic and/or screening services, procedures, medication therapy, disposition, types of health care professionals seen, and causes of injury where applicable. Data collection began in 1992 and has continued annually. Data are abstracted from 70 000 medical records of visits to 440 hospitals.

National Home and Hospice Care Survey (NHHCS). The NHHCS provides data on home health agencies and hospices and their current patients and discharges. Conducted annually since 1992, the survey was established in response to the rapid growth in the number of home health agencies and hospices throughout the USA and the need to assess the availability and utilization of these services. Personal interviews with administrators and staff provide information from approximately 1500 sample agencies. Data are collected on referral and length of service, diagnoses, number of visits, patient charges, health status, reason for discharge, and types of services provided.

National Vital Statistics System The National Vital Statistics System is responsible for the Nation's official vital statistics. These vital statistics are provided through state-operated registration systems and are based on vital records filed in state vital statistics offices. The registration of vital events—births, deaths, marriages, divorces, and fetal deaths—is a state function. NCHS cooperates with the states to develop and recommend standard forms for data collection and model procedures to ensure uniform registration of the events. The Center shares the costs incurred by the states in providing vital statistics data for national use.

Detailed annual data on births, deaths (including infant deaths), and fetal deaths are available for the USA and for states, counties, and other local areas. Monthly provisional data are available for the USA and each state. For births and deaths, data for the five OMB categories, including subgroups of the API and Hispanic population, are collected.

Linked Files of Live Birth and Infant Death Records. These research files allow researchers to explore the complex relationships between infant death and risk factors present at birth. The linked files include information from the birth certificate such as birth weight, mother's age, and prenatal care, linked to information from the death certificate for the same infant, such as cause of death and age at death. The files are birth-cohort-linked files. They are based on deaths under 1 year of age of all infants born in a calendar year. Each file contains approximately 40 000 linked records. The first annual national linked file was for the 1983 cohort under a pilot project. Beginning with the birth cohort of 1987, linked files are part of the national vital statistics system.

National Death Index. Working with state offices, NCHS established the National Death Index (NDI) as a resource to aid epidemiologists and other health and medical investigators with their mortality ascertainment activities. The NDI is a central computerized index of death record information compiled from magnetic tapes submitted by state vital statistics offices. Beginning with 1979 deaths, death records are added to the NDI file annually, approximately 10 months after the end of a particular calendar year.

The index assists investigators in determining whether persons in their studies have died and, if so, provides the names of the states in which those deaths occurred, the dates of death, and the corresponding death certificate numbers. Investigators can then arrange with the appropriate state offices to obtain copies of death certificates or specific statistical information as cause of death. The index is available to investigators solely for statistical purposes in medical and health research. It is not accessible to organizations or the general public for legal, administrative, or genealogy purposes.

THE QUALITY OF RACIAL DATA

Race is routinely used in an uncritical manner with little attention given to the underlying problems of measurement that exist for the current racial categories.⁴⁻⁷ These problems importantly affect the quality of US health data.

Measurement Error: Observer Bias

The numerator for the officially reported death rates in the USA comes from death certificates. There are reliability problems with the assessment of race that suggest that there is an acute problem of undercounting racial/ethnic status for Native Americans, APIs and Hispanics. A major source of this undercount is the discrepancy between interviewer-observed race and respondent self-report. Between 1957 and 1977 race was determined by interviewer observation in the Health Interview Survey. In 1978, the year in which the measurement of race was changed in that survey, racial information was collected both by interviewer observation and respondent self-report. Analyses of the discrepancy between these two measurement strategies revealed that 6% of persons who reported themselves as black, 29% of self-identified Asian Pacific Islanders, 62% of self-identified American Indians and 80% of persons who self-identified with an 'other' category (70% of whom were Hispanic) were classified by the interviewer as white.⁸

Respondent self-report is not an option on the death certificate, but it appears that officials who complete these forms determine racial/ethnic status based on their own judgment instead of obtaining the race of the deceased from the next of kin. A national survey of vital registrars found that they believed that only 63% of medical examiners, 50% of coroners and 47% of funeral home directors use the recommended method on relying on family members for racial information.⁹ They also indicated that funeral home directors view requesting racial information as imposing a burden on the family. Misclassification of Asian Pacific Islanders or American Indians as white would suppress the death rates for these groups. Some evidence suggests that this does, in fact, occur. Sorlie *et al.*¹⁰ compared race ascertained in a personal interview with a knowledgeable adult household member in 12 Current Population Surveys, with race recorded on the death certificate as found in the National Death Index (NDI) for the years 1979-1985. This study found high agreement for whites (99.2%) and blacks (98.2%) of self-reported racial status with racial status from death certificates. However, 26% of self-identified American Indians, 18% of Asian Pacific Islanders, and 10% of Hispanics were classified into another racial category on the death certificate. Most of these persons were classified as white.

Studies of the Indian population provide further documentation of this problem. A study in Oklahoma found that 28% of Indian infants were misclassified as another race on the death certificate.¹¹ After adjusting for this misclassification the infant mortality rate doubled from the currently reported 5.8 per 1000 to 10.4 per 1000. Similarly, another study found that only 60% of cancer patients registered with the Indian Health Service as American Indians were identified as Indians in the cancer surveillance registry.¹² This led to an underestimation of cancer incidence rates for Native Americans.

Reliability: Change in Racial Identity

Inconsistencies in reporting race/ethnicity over time is another problem. A study of a large national population found that one-third of the US population reported a different racial or ethnic status one year after their initial interview.¹³ For example, 6% of Negroes, 12% of Mexicans, 20% of Polish, 34% of Germans, and 45% of persons who said they were English, Scottish or Welsh, reported a different racial or ethnic category in 1972, compared to 1971.

The most dramatic evidence of change in self-identification comes from analyses of trends in the Indian population over time. Between 1960 and 1990 there was a six-fold increase in the American Indian population.¹³ This dramatic growth of the population cannot be explained either by biological growth or international migration. It appears to reflect a change in self-definition, with more adults of mixed ancestry identifying themselves as American Indian. This shift in self-identification into the American Indian population is more common at younger ages and does not vary by gender.^{14,15}

The degree of identification as Indian may not be very strong for many of these 'new Indians'.¹³⁻¹⁵ Most persons reporting American Indian ancestry did not report American Indian race, with 77% of persons who reported American Indian ancestry in the 1980 Census, indicating that their race was white.¹⁴ American Indian identification for this group may be optional and contextual, depending on the form of the race question, economic incentives for being American Indian in some states, reduced discrimination against American Indians, an increased willingness to self-identify as American Indians, and the increased use of self-enumeration in the Census.¹⁴ Given current rates of intermarriage of Indians with persons of other races there is likely to be continued rapid growth in the pool of persons who will be of some Indian ancestry, but for whom this ethnic identification may not be consequential. Fifty-nine percent of all married American Indians in 1990 were married to non-Indians.¹³

Definition of Racial Groups

The discussion of establishing the racial status of American Indians raises the more general problem of establishing the race of an individual whose parents are of different races. Birth certificates in the USA do not list the race of the child but they include the race of both parents. Prior to 1989 the National Center for Health Statistics used a complicated algorithm to determine the race of children whose parents belonged to different races. According to this scheme, if the father was white, the child would be given the race of the mother; but if the father was non-white the child would be assigned the race of the father. If one parent was Hawaiian then the child was Hawaiian. Thus, unlike the assignment of race for all other racial groups, a child would be white only if both parents were white. Since 1989, the National Center for Health Statistics no longer reports vital statistics by the race of the child, but reports all birth data by the race of the mother. However, the Indian Health Service continues to consider a child as Indian if either the mother or father is American Indian and there is considerable discrepancy in the publication of infant mortality rates by race of child or by race of mother. For example, in 1989 there were 39 478 American Indian births as calculated by race of mother but 49 267, as calculated by race of the child.⁹

The question of how to classify persons whose parents are of different races continues to be a hotly debated policy issue in the USA, with some groups pushing for changes in the OMB's racial standards that would include a new category of multiracial status for all persons whose parents come from more than one of the four official racial groups. The current trend of interracial marriage suggests that this question will apply to an ever-increasing proportion of the population. Twenty-five percent to 44% of Hispanics marry non-Hispanics and from 25% to 50% of API subgroups marry persons of other

rac¹⁶ Rates of black-white intermarriage are considerably lower but they increased from 2% in 1970 to 6% in 1990. The states of Ohio and Illinois now require that the multiracial category must be added to school forms while Georgia and Michigan requires it on written forms used by state agencies.

Researchers have not given systematic attention to the extent to which the health profile of persons of mixed racial parentage differs from that of the standard racial categories. One recent study suggests that this association may be complex and that any attempt to assess multiracial status should include assessment of the race of both parents. Collins and David¹⁷ found that infants born to black mothers and white fathers had a higher rate of low birth weight than those born to white mothers and black fathers.

Definitional problems are not limited to multiracial status. The classification of the entire population into racial groups is neither simple nor straightforward. These problems are readily evident for the American Indian population. Indian tribes do not agree on who is American Indian with some using a strict definition based on blood quantum level, while others require identification with Indian culture or participation in tribal affairs. In the USA, a national debate continues over how race should be conceptualized and measured.¹⁸ There is disagreement over the optimal terminology to be used for particular racial groups, whether new ethnic categories should be added (for persons from the Middle East or the Cape Verde islands), and whether Hispanic should be a racial or an ethnic category.

The categories of race and ethnicity are often assessed differently by various health agencies. The Centers for Disease Control and Prevention uses one question to capture the OMB's five required racial categories in its national notifiable diseases surveillance system. However, many health surveys and other Federal data collection systems use one question to assess race and a separate one to assess Hispanic origin. Similarly, although the standard birth and death certificates were revised in 1989 to include Hispanic identifiers, the wording of the question varies from state to state and the data on Hispanic ethnicity is not completely reported in all states.¹⁹

These differences are important because the size of a racial/ethnic population depends on the wording of the question. In the 1980 Census 26.5 million Americans self-identified as 'black or Negro', but only 21 million indicated that they were of Afro-American ancestry.⁴ Similarly, there were 1.5 million American Indians based on answers to the race question in the 1980 Census, but 6.8 million based on responses to the ethnic ancestry question.²⁰

Respondents also vary in their preferred term for self-identification. A recent national study of over 60 000 adults found that members of racial groups are divided over preferred terminology.²¹ Fifty-eight percent of Hispanics preferred 'Hispanic' (12% prefer 'Latino'), 62% of whites preferred 'white' (17% prefer 'caucasian'), 44% of blacks preferred 'black' (28% prefer 'African American'), and 50% of American Indians prefer the term 'American Indian' (37% prefer 'Native American'). In an effort to respect individual dignity, researchers should use the most preferred terms for each group interchangeably (e.g. black or African American, Hispanic or Latino).

The implementation of the OMB's Directive 15 frequently violates an important principle of classification, namely, the creation of a set of exhaustive and mutually exclusive groups. Valid statistical tests are based on the assumption that the various categories in the classification system are independent samples. This assumption is not met for much of the routine reporting of health data in the USA (that provides information for non-Hispanic whites, non-Hispanic blacks, Hispanics and APIs). There are black Hispanics, Asian Hispanics, American Indian Hispanics and white Hispanics. Del Pinal²² has shown that the overlap of race with the Hispanic category affects the patterns of racial/ethnic differences not only for Hispanics but for the other racial categories as well.

Census Undercount

Another problem affecting the quality of health data in the USA is Census undercount. Census data are used to calculate the denominators for mortality rates. They are also used to construct sampling frames and adjust for nonresponse in population-based epidemiologic studies. The use of a denominator that is undercounted inflates the obtained rate in exact proportion to the undercount in the denominator. Thus, all rates of health events that use census data as denominators are overestimated by the same percentage as the population undercount in the denominator.

For the last several decades the USA Census Bureau has been evaluating the extent of undercount by means of demographic analysis. This strategy produces estimates of the population based on administrative data and demographic trends. Consistently, these analyses reveal that, while the overall undercount for the US population is small, it is larger for blacks than for whites, and despite a steady decline in the undercount rate for blacks between 1940 and 1980, there was an upward trend between 1980 and 1990.²³ The undercount rate for the overall population does not importantly distort health data, but the undercount rate varies considerably for some demographic subgroups.

In 1990, the overall undercount was 1.8% for the US population and 5.7% for the black population. However, census undercount was dramatically higher for black males (8.5%) than for black females (3%), and varied by age such that there was a net census undercount of 11–13% for all of the 10-year age categories of black males between the ages of 25 and 64.²⁴ Demographic analysis estimates are available only at the national level and it is likely that the omission of black males from households (the major cause of the undercount of blacks) varies by geographic area. Estimates of undercount based on demographic analyses are only as good as the underlying assumptions and concerns have been raised about the extent to which the demographic analysis methods are becoming less reliable over time.²⁵

The evaluation of the undercount problem by the Census Bureau has focused heavily on the black and white population. However, there is reason to believe that census undercount may also be a significant problem for some of the other racial populations. For the 1990 Census, in addition to demographic analysis, the Census conducted a Post Census Enumeration Survey (PES) in which undercount was estimated on a case by case matching of Census records with those obtained in the survey of 165 000 households. According to the PES, the undercount rates for Hispanics (5%) and reservation Indians (12.2%) were even higher than the rate for blacks (4.6%), but the extent to which the undercount for these groups is concentrated in particular age and/or gender groups is not known.²⁶ The PES undercount was 0.7% for non-Hispanic whites and 2.4% for Asian and Pacific Islanders.

CURRENT ISSUES IN THE MEASUREMENT OF RACE/ETHNICITY

Conceptualization of Race

Traditional explanations for health status differences between the races in the USA have focused on biological differences between racial populations. Nineteenth-century medical research, for example, attempted to document that blacks were biologically inferior to whites and therefore more susceptible to a host of illnesses.²⁷ Most current research on racial differences has abandoned blatant racist ideology but much of it still assumes that racial variations in disease are due to underlying differences in biology. The biological approach views racial taxonomies as meaningful classifications of genetic differences between human population groups. It assumes that race is a valid biological category, that the genes which determine race also determine the number and types of

health problems that an individual will have.²⁸ The available scientific evidence suggests that race is a social and not a biological category. First, the concept of race developed long before modern valid scientific theories of genetics existed. Historically, race was a useful construct, not only for classifying human variation, but also for providing a rationale for the exploitation of groups that were regarded as inferior.²⁹ Second, the extant racial categories do not represent biological distinctiveness. There is more genetic variation within races than between them. Irrespective of geographic origin or race, all human beings are identical for 75% of known genetic factors.³⁰ In addition, some 95% of human genetic variation exists within racial groups with relatively small and isolated populations such as Eskimos and Australian Aborigines contributing most of the between-group variation.³¹ Thus, our current racial categories are more alike than different in terms of biological characteristics and genetics and there is no specific scientific criteria to unambiguously distinguish different racial groups.

Moreover, single gene disorders account for only a small part of racial differences in health. Sickle cell anemia, for example, occurs more frequently in African Americans than the rest of the population. However, it accounts for only three-tenths of 1% of the total number of excess deaths in the black population and is thus not a major cause of the higher rates of disease for African Americans.³² Sickle cell anemia appears to have been a protective biological adaptation to environmental conditions. It is not limited to African Americans but occurs at higher rates for persons who originate in regions of the world where malaria was endemic.³³

Race and SES

Since the discrediting of biological explanations, researchers have been giving increasing attention to the role of social class or SES as a determinant of racial differences in health. Race is strongly associated with SES and many researchers view race as a proxy for SES. Table 2 presents selected SES characteristics for the five OMB categories.³⁴ The overall patterns for racial/ethnic groups in Table 2 mask the considerable variation that exists within each of the categories.

The data on educational attainment reveal that blacks, Hispanics, and American Indians have considerably lower rates of educational attainment than whites. Rates of high school completion for Asian and Pacific Islanders are comparable to those of whites with the APIs having even higher levels of college graduation than whites. The unemployment data show a similar pattern with APIs and whites having the lowest rates and with the rates being considerably higher for American Indians, blacks and Hispanics in that order. However, other data reveal that the unemployment rates for some Asian subgroups are also high. The 1980 unemployment rates were 20% for the Hmong, 15% for Laotians, 11% for Cambodians and 10% for Samoans.³⁵ The median household income for American Indians (\$19,865) is very similar to that of blacks (\$19,758) and both groups have income levels that are considerably lower than that of the white population (\$31,435). The median household level for Hispanics is higher than that of blacks and American Indians but lower than that of whites, while the Asian population has the highest level of median household income in the USA. However, Asian families are larger and have more earners per family than the total population. Thus, the 1990 per capita income of Asians (\$13,420) was lower than that of whites (\$15,270).³⁶ The data on poverty also follow the now familiar pattern. The highest rates of poverty are found for American Indians and blacks. These groups have rates that are about three times that of whites. Hispanics have rates that are substantially higher than whites, but lower than those of blacks and American Indians. The rates for the API category are slightly higher than that of whites.

TABLE 2. *Selected economic profiles for the USA, 1990 Census*

	American Indian and Alaskan Native	White	Black	Asian and Pacific Islander	Hispanic
Educational attainment (persons 25 years and older):					
Percent less than 9th grade	14.2	8.9	13.8	12.9	30.7
Percent high school grad or higher	65.3	77.9	63.1	77.5	49.8
Percent bachelor's degree or higher	8.9	21.5	11.4	36.6	9.2
Employment status by sex (persons 16 years and older):					
Percent unemployed, males	16.2	5.3	13.7	5.1	9.8
Percent unemployed, females	13.5	5.0	12.2	5.5	11.2
Household income in 1989:					
Median household income	\$19,865	\$31,435	\$19,758	\$36,784	\$24,156
Percent below the poverty level by age:					
All ages	31.7	9.8	29.5	14.1	25.3
Under 5 years	43.3	13.8	44.0	17.5	33.4

Source: Department of Health and Human Services.

SES and Racial Differences in Health

Given this strong association between SES and race, it is widely recognized that racial differences must be controlled for SES. Adjusting racial (black/white) disparities in health for SES sometimes eliminates but always substantially reduces these differences.³⁷⁻⁴⁰ However, it is frequently found that within each level of SES blacks still have worse health status than whites. This pattern suggests that although most of the racial differences in health are accounted for by SES, race has an effect on health that is independent of SES. That is, while there is considerable overlap between race and SES, race reflects more than SES and fully understanding racial differences in health will require researchers to move beyond the traditional approaches.

Several researchers have emphasized that the failure of the traditionally utilized SES indicators to completely explain racial differences in health reflects the interactive and incremental role of racism as a determinant of health. The construct racism incorporates ideologies of superiority, negative attitudes and beliefs toward racial and ethnic out-groups, and differential treatment of members of these groups by both individuals and societal institutions.⁴⁰

Racism can affect health by giving rise to racial discrimination at the individual and institutional level. The former is an important but neglected stressor that can lead to adverse changes in health status while the latter can result in the inequitable distribution of desirable institutional resources including medical ones. However, racism is causally prior to SES and exerts its most profound impact by transforming SES such that an equivalent value on a traditional SES measure represents important differences in social and economic circumstances for persons belonging to different racial groups.⁴¹ For example, college-educated blacks are about four times more likely to experience unemployment than their white peers.⁴² Even after adjustment for job experience and training, blacks are more likely than whites to be exposed to occupational hazards and carcinogens at work.⁴³ Table 2 had earlier noted large racial differences in income, but income disparities understate racial differences in household economic resources. Racial differences in wealth are larger than those for income. Thus, compared to whites at equivalent levels of income, blacks and Latinos have substantially less economic security and are less able to cushion a shortfall of income.

Heterogeneity of racial populations An emerging issue in the assessment of racial and ethnic status is the need to collect sufficient information to examine the variation within each of the standard categories. The American Indian (or Native American) population is characterized by considerable diversity. There are more than 500 federally recognized tribes and entities. Death rates for American Indians also vary considerably from state to state with rates being higher in states that have a larger concentration of Indians. Moreover, considerable tribal-specific variation often exists within a specific state. For example, within the state of New Mexico there are large tribal differences in prenatal care, low birth weight and infant mortality.⁴⁴

Although the API population in the USA is geographically concentrated, with almost 80% of all APIs residing in only 10 states, the API category lumps together persons coming from 28 Asian countries and 25 Pacific Island cultures.³⁶ Each of these subgroups has its own distinctive history, culture and language. Not surprisingly, an overall value on a health status indicator for the API population hides the considerable heterogeneity that exists for subgroups within that population. For example, the API population in California has death rates of homicide and legal intervention for 15-24-year-olds that is 17 per 100 000, but the rates range from 6 for Chinese Americans and 13 for Japanese Americans, to 54 for Samoans and 73 for the other Pacific Islander category.⁴⁵ Similarly, while the API population has the lowest death rates of any racial group in the USA, Native Hawaiians have the highest death rate due to heart disease of

any racial group in the USA,⁴⁶ and the rate of liver cancer for Chinese Americans is four times higher than that of the white population.³⁶ Similarly, while most Hispanics have a common language, religion, and various traditions, the timing of immigration and the incorporation experience in the USA have varied for the more than 25 national origin groups that make up the Hispanic group, such that each group is distinctive.⁴⁷

Researchers have also given inadequate attention to the variations within both the black and white population. The black population is characterized by cultural and ethnic heterogeneity that is predictive of variations in health status.⁴¹ There is considerable ethnic variation in the white population but little recent research attention has been given to examining the extent to which these differences predict variations in health status.

Comparative Analysis. The 1991 Census of England and Wales was the first to utilize the question on ethnicity.⁴⁸ Seven preassigned codes were used in the census. These were: white, black Caribbeans, black Africans, Indians, Pakistanis, Bangladeshis, and Chinese. An additional 28 categories were developed in the census based on write-in responses under the 'Black other' and 'Any other ethnic group' categories. Many summary classifications used in the census data includes the seven preassigned codes plus three additional codes: Black others, Other Asian, and Other. In addition to the question on ancestry, a question on country of birth was utilized and over 102 categories were developed. The question on country of birth has been utilized in previous censuses.

Several recent health surveys in the UK have also included measures of ethnic group membership. The Health Survey of England which began in 1991 and focuses heavily on cardiovascular disease and risk factors for cardiovascular disease has included ethnic group membership in the survey. Similarly, the General Household Survey, a large annual omnibus survey, of about 15 000 households per year has included data on ethnicity and country of birth in recent years. The survey includes measures of self-reported illness, both chronic and acute, as well as of risk factors and socioeconomic status. The OPCS national morbidity studies carried out around the time of the censuses are a major source of information of ambulatory care utilization in the UK. In 1991, data were collected on ethnicity and country of birth so it is now possible to generate rates of utilization of general practitioners by racial and ethnic status. The major limitation of this study is that it involves only self-selected general practitioners and there are questions about both generalizability, as well as the recoverage of the racial and ethnic populations.

Data on patient care in the UK is available through the Hospital Episode Statistics system. This system collects data on all inpatient and outpatient hospital visits. In 1995, ethnic categories were included in this data system. It will thus become possible to examine rates of particular diseases, surgical interventions, severity, etc. by ethnic group status. Ethnic status is not collected on death certificates and it would require an Act of Parliament for the inclusion. However, country of birth has been collected on death certificates since 1969 so that it is easy to examine mortality rates for first generation immigrants. Similar to the USA, there is a linked infant birth and death file in the UK but unlike the USA, ethnicity is not collected at the time of registration of birth.

RECOMMENDATIONS

1. There is a continuing need for uniform assessment of race and ethnicity by government-administered health data collection systems as well as by the wider research community.
2. Given the heterogeneity of racial and ethnic populations, there is also a critical need for the inclusion of identifiers for subgroups of the API and Hispanic populations on all surveys and forms in the USA. The 10 categories in the UK Census appears to

capture the key dimensions of ethnic heterogeneity given the size of these ethnic minority populations.

3. Racial/ethnic data should be routinely utilized in the design, implementation and evaluation of health studies and health programs. The availability of adequate data, especially morbidity data, for American Indians, Hispanics, and APIs is still a major problem. Because of the relatively small sizes of some of these population groups and their geographic distribution, standard sampling strategies for national populations do not yield adequate sample size to provide reliable estimates for the distribution of disease in these groups or to explore heterogeneity within a given racial group. Surveys focused on a particular geographic area with a high concentration of a racial subgroup as opposed to national ones are necessary to provide data for these groups. Combining multiple years of data in ongoing surveys is another useful strategy for obtaining health information for small population groups. This latter strategy can also be useful in the UK context in population-based health surveys.
4. Translate questionnaires and ensure that measurement instruments are culturally appropriate. Health researchers must also give greater attention to translating study instruments for persons who have limited proficiency in the English language. These persons are more likely to be members of racial minority populations. For example, in 1990, while only 8% of the total US population was foreign born, 74% of APIs were foreign born.³⁶ Currently, major federal health surveys, such as the Health Interview Survey, do not routinely translate the questionnaire into other languages. In addition to translating the survey instruments, researchers must also ensure that their new instruments meet the tests of conceptual, scale and norm equivalence.³⁵ Conceptual equivalence refers to similarities in the meanings of the concepts used in the assessment. Scale equivalence is the use of questionnaire items that are familiar to all groups, while norm equivalence ensures that the norms developed for the targeted group is appropriate and not arbitrarily assigned from another.
5. Build communication mechanisms with racial/ethnic communities to ensure that they receive findings from current studies and have input in future research and interventions.
6. Periodically monitor and update the current measures of race and ethnicity. Data collection forms should be revised to ensure consistent classification across data systems.
7. Whenever racial/ethnic data is reported, give more attention to interpretation: always indicate why race/ethnicity is being used, the limitations of racial/ethnic data, and how findings should be interpreted. The presentation of data on racial differences should routinely stratify them by SES within racial groups. Failure to do this may mis-specify complex health risks and even lead to harmful social stereotypes and consequences.
8. Move from studies of race and health to studies that identify the specific factors linked to race that affect health. Whenever feasible, additional information that captures these characteristics should be collected. This will include the assessment of SES, acculturation, and economic and noneconomic aspects of discrimination.⁴⁹ There are limited opportunities to collect additional information in the vital statistics system and in record-based surveys. However, even in these contexts, years of formal education, nativity status and years since migration can be ascertained.

REFERENCES

1. Martin E, DeMaio TJ & Campanelli PC. Context effects for census measures of race and Hispanic origin. *Public Opinion Quarterly* 1990; 54: 551-566.
2. Anderson M & Feinberg SE. Black, white, and shades of gray (and brown and yellow). *Chance* 1995; 8: 15-18.

3. Omi M & Winant H. *Racial Formation in the United States: From the 1960s to the 1980s*. New York: Routledge, 1986.
4. Hahn RA. The state of federal health statistics on racial and ethnic groups. *Journal of the American Medical Association* 1992; 267: 268-79.
5. Williams DR. The concept of race in health services research: 1966-1990. *Health Services Research* 1994; 29: 261-74.
6. LaVeist TA. Beyond dummy variables and sample selection: what health services researchers ought to know about race as a variable. *Health Services Research* 1994; 29: 1-16.
7. Jones CP, LaVeist TA & Lillie-Blanton M. Race in the epidemiologic literature: an examination of the American Journal of Epidemiology, 1921-1990. *American Journal of Epidemiology* 1991; 134: 1079-84.
8. Massey JT. A comparison of interviewer observed race and respondent reported race in the National Health Interview Survey. *1980 Proceedings of the Social Statistics Section*. American Statistical Association, Washington, DC, 1980: 425-28.
9. Scott S & Suagee M. *Enhancing Health Statistics for American Indian and Alaskan Native Communities: An Agenda for Action. Report to the National Center for Health Statistics*. St Paul, MN: American Indian Health Care Association, 1992.
10. Sorlie PD, Backlund E, Johnson NJ & Rogot E. Mortality by Hispanic status in the United States. *Journal of the American Medical Association* 1993; 270: 2464-2468.
11. Kennedy RD & Deapen RD. Differences between Oklahoma and Indian infant mortality and other races. *Public Health Reports* 1991; 106: 97-99.
12. Frost F, Taylor V & Fries E. Racial misclassification of Native Americans in a surveillance, epidemiology, and end results cancer registry. *Journal of the National Cancer Institute* 1992; 84: 957-962.
13. Eschbach K. The enduring and vanishing American Indian: American Indian population growth and intermarriage in 1990. *Ethnic Racial Studies* 1995; 18: 89-108.
13. Johnson CE. *Consistency of Reporting Ethnic Origin in the Current Population Survey*. U.S. Department of Commerce Tech. Pap. No. 31. Washington, DC: Bureau of the Census, 1974.
14. Passel JS & Berman PA. Quality of 1980 Census data for American Indians. *Social Biology* 1986; 33: 163-82.
15. Harris D. The 1990 Census count of American Indians: what do the numbers really mean? *Social Science Quarterly* 1994; 75: 580-93.
16. Rumbaut RG. The crucible within: ethnic identity, self-esteem, and segmented assimilation among children of immigrants. *International Migration Review* 1994; 28: 748-94.
17. Collins JW & David RJ. Race and birthweight in biracial infants. *American Journal of Public Health* 1993; 83: 1125-29.
18. Evinger S. How shall we measure our nation's diversity? *Chance* 1995; 8: 7-14.
19. US Public Health Task Force on Minority Health Data. *Improving Minority Health Statistics*. Report 715-025. Washington, DC: USGPO, 1992.
20. Snipp M. *American Indians: The First of this Land*. New York: Russell Sage Foundation, 1989.
21. Tucker H, McKay R, Kojetin B, Harrison R, de la Puente M, Stinson L & Robison E. Testing methods of collecting racial and ethnic information: results of the current population survey supplement on race and ethnicity. *Bureau of Labor Statistical Notes* 1996; 40: 1-149.
22. del Pinal JH. *Exploring alternative race-ethnic comparison groups in current population surveys*. U.S. Bureau of the Census, *Current Population Reports, Series P23-182*. Washington, DC: USGPO, 1992.
23. Robinson JG, Bashir A, Prithwis DG & Woodrow KA. Estimation of population coverage in the 1990 United States Census based on demographic analysis. *Journal of the American Statistical Association* 1993; 88: 1047-57.
24. National Center for Health Statistics. *Vital Statistics of the United States, 1990, Vol. 11, mortality, part A*. Washington, DC: Public Health Service, 1994.
25. Notes and Comments. Census undercount and the quality of health data for racial and ethnic populations. *Ethnicity and Disease* 1994; 4: 98-100.
26. Hogan H. The post-enumeration survey: operations and results. *Journal of the American Statistical Association* 1993; 88: 1047-57.
27. Krieger N. Shades of differences: theoretical underpinnings of the medical controversy on black/white differences in the United States, 1830-1870. *International Journal of Health Services* 1987; 17: 259-78.
28. Krieger N & Bassett M. The health of black folk: disease, class, and ideology in science. *Monthly Review* 1986; 38(3): 74-85.
29. Montagu A. *The Concept of Race*. New York: Free Press, 1965.
30. Lewontin R. *Human Diversity*. New York: Scientific American Books.
31. Lewontin R. *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press, 1974.

32. Cooper RS & David R. The biological concept of race and its application to public health and epidemiology. *Journal of Health and Politics, Policy, and Law* 1986; 11: 97-116.
33. Polednak AP. *Racial and Ethnic Differences in Disease*. New York: Oxford University Press, 1989.
34. National Center for Health Statistics. *Trends in Indian Health—1993*. US Dept. of Health and Human Services, Rockville, MD: Indian Health Service, 1993.
35. Takeuchi DT & Young KNJ. Overview of Asian and Pacific Islander Americans. In: Zane NWS, Takeuchi DT & Young KNJ, eds. *Confronting Critical Health Issues of Asian and Pacific Islander Americans*. Thousand Oaks, CA: Sage, 1994: 3-21.
36. Lin-Fu JS. Asian and Pacific Islander Americans: an overview of demographic characteristics and health care issues. *Asian and Pacific Islander Journal of Health* 1993; 1: 20-36.
37. Krieger N, Rowley DL, Herman AA, Avery B & Phillips MT. Racism, sexism, and social class: implications for studies of health, disease, and well-being. *American Journal of Preventive Medicine* 1993; 9 (6 suppl): 82-122.
38. Cooper RS. Health and the social status of blacks in the United States. *Annals of Epidemiology* 1993; 3: 137-44.
39. Krieger N & Fee E. Social class: the missing link in U.S. health data. *International Journal of Health Services* 1994; 24: 25-44.
40. Williams DR & Collins C. U.S. socioeconomic and racial differences in health. *Annual Review of Sociology* 1995; 21: 349-86.
41. Williams DR, Lavizzo-Mourey R & Warren RC. The concept of race and health status in America. *Public Health Reports* 1994; 109(1): 26-41.
42. Wilhelm SM. Economic demise of blacks in America: a prelude to genocide? *Journal of Black Studies* 1987; 17: 201-54.
43. Robinson J. Racial inequality and the probability of occupation-related injury or illness. *Milbank Memorial Fund Quarterly* 1984; 62: 567-90.
44. Halasan C *et al.* 1990-1991 *New Mexico tribe-specific vital statistics*. State of New Mexico, Department of Health, Public Health Division, Bureau of Vital Records and Health Statistics, Santa Fe, NM, 1992.
45. Suh D. Cooperative agreements to advance the understanding of the health of Asian and Pacific Islander Americans. In: Centers for Disease Control and Prevention, *Proceedings of the 1993 Public Health Conference on Records and Statistics*. DHHS Publication No. (PHS) 94-1214. Hyattsville, MD: National Center for Health Statistics, 1993: 352-56.
46. Chen MS. A 1993 status report on the health status of Asian Pacific Islander Americans: comparisons with Healthy People 2000 objectives. *Asian American Pacific Islander Journal of Health* 1993; 1: 37-55.
47. Massey DS & Denton NA. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press, 1993.
48. Balarajan R. Information on ethnicity and health in England and Wales. Paper prepared for Health Gain for Black and Minority Ethnic Communities: A US/UK Conference. 4(1) 5-15.
49. Williams DR. Race and health: basic questions, emerging directions. *Annals of Epidemiology* 7(5), 322-333.