CLASSIFICATION OF RACE AND ETHNICITY:  
Implications for Public Health*

Vickie M. Mays,1 Ninez A. Ponce,2 Donna L. Washington,3 and Susan D. Cochran4

1Department of Psychology, University of California, Los Angeles, Box 951563, Los Angeles, California 90095-1563; email: mays@ucla.edu
2Department of Health Services, University of California, Los Angeles, School of Public Health, Los Angeles, California 90095-1772; email: nponce@ucla.edu
3Department of Medicine, Veterans Affairs, Greater Los Angeles Healthcare System, 11301 Wilshire Boulevard, 111G, Room 3242, Los Angeles, California 90073; email: donna.washington@med.va.gov
4Department of Epidemiology, University of California, Los Angeles, School of Public Health, California 90095-1772; email: cochran@ucla.edu

Key Words  classification bias, public health statistics, cultural competence

Abstract  Emerging methods in the measurement of race and ethnicity have important implications for the field of public health. Traditionally, information on race and/or ethnicity has been integral to our understanding of the health issues affecting the U.S. population. We review some of the complexities created by new classification approaches made possible by the inclusion of multiple-race assessment in the U.S. Census and large health surveys. We discuss the importance of these classification decisions in understanding racial/ethnic health and health care access disparities. The trend toward increasing racial and ethnic diversity in the United States will put further pressure on the public health industry to develop consistent and useful approaches to racial/ethnic classifications.

INTRODUCTION

The collection of data by the federal government to classify individuals by their race has a long and somewhat contentious history in the United States (3). In recent years, some have challenged the methods or purposes of classifying race on the grounds that race is a social construct and not an essentialistic feature of human beings that can be reliably measured in order to draw meaningful conclusions (40, 57, 126, 127). Each of us writing this paper confesses at the start that we share, along with a number of social scientists, the perspective that race, as currently

*The U.S. Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.
measured in most health surveys, reflects a social construct in the minds of Americans that is imprecisely mapped to the differences that exist among us. However, we also recognize that both federal and state governments, using mechanisms such as the Census or vital health statistics, classify individuals by their race and ethnicity in the hope of obtaining useful information to improve the public health. In what follows, we sidestep the bulk of the essentialist-social constructivist debate to focus instead on practical, applied problems in the classification of race and ethnicity in federal health statistics, as well as the implication of these classifications for public health.

There are several reasons why public health is struggling with the problem of measuring race and ethnicity. In recent years, newly emerging trends in demography, science, and public policy have once again brought to the forefront the difficulties of measuring constructs that are far more complex than they might appear at first blush. Changing demographics in the United States make categories from even a half century ago inadequate, whatever one’s perspective on the nature of race and ethnicity (1, 12, 35, 39, 69, 89, 109, 112, 119, 123, 129). Further, this trend is likely to accelerate with the current pattern of immigration to the United States, reductions in social barriers against racial or ethnic intermarriages, and increasing racial and ethnic heterogeneity in cities, job sites, and neighborhoods. An increasing percentage of the American population can now trace its roots to multiracial or multiethnic sources. At the same time, rapid gains in science, such as the Human Genome Project, promise eventual health benefits from knowing one’s genetic make up, including information about diseases that are associated with racial or geographic origins (86, 102, 118). To the extent that measurement of current racial or ethnic status can provide accurate information concerning genetic liabilities, then the health benefits of these new findings might be further maximized. Finally, there is a growing recognition that current methods of assessing race/ethnicity are insufficiently precise for the needs of researchers. Community groups too have come to recognize the usefulness of health statistics in addressing local health needs, but this information must be specific and appropriate to the neighborhoods in which they live.

TRADITIONAL APPROACHES TO RACE AND ETHNICITY CLASSIFICATION IN PUBLIC HEALTH

The Use of Race and Ethnicity Information in Public Health

Traditionally, data are collected on race and ethnicity to accomplish several overlapping purposes in public health.

TO DESCRIBE VITAL AND HEALTH STATISTICS Estimates of vital statistics and the burden of morbidity and mortality in a particular population are of great use to government, in general, and to the field of public health, in particular, for purposes of planning, tracking the needs of citizens, and identifying modifiable health risks. Doing so is dependent on both well-articulated definitions of the source population
in the abstract and the ability to accurately classify elements, both case and non-case, as either in the source population or not (105). Counts of the source population are used as denominators in many statistics, including health statistics. Moreover, cases (often a component in the numerator of health statistics) need to be accurately paired with their source population. In order to generate estimates that are especially useful to health planning, source populations are often narrowly drawn (e.g., women of childbearing age, African Americans). Misclassifications can result in misleading estimates of both health and disease burdens.

**AS A RISK INDICATOR FOR HEALTH OUTCOMES** Individual variations in risk for selected disorders, response to medications, or other interventions, and health outcomes have been the subject of much study (55, 120, 126). Variations associated with race or ethnicity may stem from differences in biology (132), behavior (94), and/or exposure to environmental factors (117). Similarly, variations in health outcomes have been linked to biologic determinants, social determinants, (5, 121) and their interaction (53). Although race or ethnicity, per se, is not causal, it may still function as a risk indicator providing some reduction in uncertainty about the likelihood of morbidity or mortality (105).

**TO IMPROVE THE DELIVERY OF HEALTH SERVICES** Knowledge of the race/ethnicity of an individual or of the racial/ethnic distribution of a population has the potential to allow for more effective tailoring of health services delivery. This may occur at all levels of health care intervention. The HHS Initiative on Eliminating Racial/Ethnic Disparities in Health in the United States is an example of how racial/ethnic categorizations are being incorporated into public health objectives, program planning, and interventions on a national level (116).

**AS A MARKER OF UNMEASURED BIOLOGICAL DIFFERENCES** Information about race or ethnicity, gathered by self-report or other public health methods of record keeping, is often used in the field as a marker for biological substrates that are not directly measured. One’s confidence in its accuracy is perhaps dependent on whether or not race is viewed as a biological entity (70). Some biological differences among individuals reflect, in part, the adaptation of human groups to environmental conditions that lead to greater prevalence of some diseases due to vulnerabilities inadvertently shaped by exposure to other diseases. One such example is sickle-cell anemia, thought to reflect a protective adaptation to malaria. Prior to large migrations and rapid intermingling of human populations, such as what has occurred in the past several hundred years, characteristics developed over many thousands of years, which today are labeled as markers of race or ethnicity, were presumably more tightly associated with some genetic patterns. Thus, race/ethnicity is still highly correlated with the occurrence of a few, specific chronic diseases including sickle-cell anemia among African Americans or Tay-Sachs disease among Ashkenazi Jews. In these instances, both are single-gene disorders. However, genetic factors account for only a minimal amount of the observed racial/ethnic variations in health (10, 20, 70, 75).
Efforts to link race/ethnicity classifications of individuals to biological typologies are more commonly not successful. For example, pharmacogenetics research has explored racial/ethnicity-linked variations in single nucleotide polymorphism—the sites where DNA sequence differs among individuals. Single nucleotide polymorphism profiles are also associated with response to certain medications. The advantage of linking these to observable or measurable racial/ethnic categories would be ultimately to improve patient care. However, in a study that compared 23 markers for such genes among 354 people representing 8 classically defined races, the genetic markers were found to form 4 natural groupings that did not correspond to any of the phenotypically defined categories (131). Indeed, despite the current focus in genomic research on minute differences between individuals and among groups, announcements from the Human Genome Project on the mapping and sequencing of the human genome highlight the finding that the human population shares 99.9% of its DNA (70, 86, 102, 118).

Cooper, Williams, Krieger, and LaVeist have written extensively about the limitations of a biological model in etiologic thinking in public health (20, 21, 60, 61, 67, 130). To better understand the basis for observed racial/ethnic variations in health, each supports the collection of additional data that capture the specific factors that contribute to group differences (58–60, 64, 66, 67, 127, 128). In general, however, between-group differences among phenotypic classified racial/ethnic groups are much smaller than the diversity that exists within groups, suggesting that biologic contributions will be, in most instances, minimal as compared to other possible explanatory factors (38, 70).

**AS A PROXY FOR UNMEASURED SOCIAL FACTORS** Cultural, social, and environmental influences also contribute to the considerable heterogeneity in the distribution of disease and risk factors for disease. These influences vary by racial/ethnic background. Because information on race/ethnicity is commonly available in health data, this information can then be used as a proxy for unmeasured social factors. The extent to which this strategy is effective varies substantially and is dependent, of course, on the strength of association between the proxy (e.g., race or ethnicity) and the unmeasured construct (10, 70). For example, cultural norms are known to influence individual levels of behavior, such as dietary practices, tobacco and alcohol use, or responses to stressful events (67, 94). Many of these behaviors are associated with both race or ethnicity and health outcomes.

This association between cultural, social, and environmentally influenced behaviors and phenotypic race or ethnicity provides a partial explanation for observed racial/ethnic disparities in health. For example, Otten and colleagues found that six well-established risk factors (tobacco use, hypertension, hyperlipidemia, alcohol intake, excess weight, and diabetes mellitus) accounted for 31% of the excess mortality between Black and White adults in the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study (94).

Social factors, including income, education, insurance status, and having a regular source of health care, are also well-established determinants of health.
These, too, differ among racial/ethnic groups (17, 76, 80, 88, 95, 125). But unlike race or ethnicity, these cultural, social, and environmental factors are mutable and potentially amenable to public health intervention. Among African Americans, for example, differences in delays in seeking emergency care for acute chest pain have been found to be related to socioeconomic status (32) in part, though not completely (101).

Measurement of Race and Ethnicity: Old and New Standards

OMB DIRECTIVE 15 1977 The measurement of race by the federal government has been ongoing in conjunction with the decennial census since 1790 (see Table 1). Over the years, the Census Bureau has used various approaches and criteria to classify the U.S. population including methods dependent on national origin, tribal affiliations, and physical characteristics (70). The terms mullatto, quadroon, and octaroon emerged from nineteenth-century classifications in which individuals were categorized by their percentage of Black African–based ancestry. This “one drop” of Black African ancestry for determining categorization reflects an early method of operationalizing a biologic classification scheme (10, 70). Throughout the twentieth century, as illustrated in Table 1, different templates have been used to categorize racial differences. For example Jews early in the century were classified as non-White but later “deracialized” (70) and counted as White. In 1977, the Office of Budget and Management (OMB) released Statistical Policy Directive No. 15 Race and Ethnic Standards for Federal Statistics and Administrative Reporting (90). This directive was one of the first attempts at the federal level to create standards and consistency across all of the data gathering activities of the federal government. This was pushed by the need for data to assess compliance with the Civil Rights Act of 1964 and the Voting Rights Act of 1965 and their various amendments and interpretations that continued into the early 1970s (31, 70). The 1977 Directive specified that data on race be reported for what was viewed as four mutually exclusive single-race categories: White, Black, American Indian/Alaska Native, and Asian and Pacific Islander. Ethnicity was specified to be collected in a minimal fashion: as persons of Hispanic or not of Hispanic origin.

OMB STANDARDS FOR MAINTAINING, COLLECTING, AND PRESENTING FEDERAL DATA ON RACE AND ETHNICITY As the population of the United States grew more diverse, it became increasingly apparent that the 1977 classification did not adequately capture the way individuals thought about themselves. The occasion of the 1990 Census was a significant factor in the changes recommended by OMB in 1997 (35a). One third of the population growth between the 1980 and 1990 counts were due to the arrival of immigrants whose origins confounded categories of “Hispanic” and “Asian and Pacific Islander.” Table 1 shows the evolution of the classification scheme of the Census since 1850. Note the expanding categories of Asian and Pacific Islander, and American Indian and Alaskan Natives over the past 50 years (31, 70). By the 1990 Census, researchers began to hone in on the
[table 1]

<table>
<thead>
<tr>
<th>Year</th>
<th>White</th>
<th>Black/ Negro</th>
<th>Native peoples</th>
<th>Chinese</th>
<th>Japanese</th>
<th>Other Asian or Pacific Islander</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1850</td>
<td></td>
<td>Black, mulatto</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1860</td>
<td></td>
<td>Black, mulatto</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1870</td>
<td>White</td>
<td>Black, mulatto</td>
<td>Indian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td>White</td>
<td>Black, mulatto</td>
<td>Indian</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1890</td>
<td>White</td>
<td>Black, mulatto,</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1900</td>
<td>White</td>
<td>Black</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1910</td>
<td>White</td>
<td>Black, mulatto</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td>White</td>
<td>Black, mulatto</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Hindu, Filipino,</td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td>White</td>
<td>Negro</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Filipino,</td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td>White</td>
<td>Negro</td>
<td>Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Filipino,</td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>White</td>
<td>Negro</td>
<td>American Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Filipino,</td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>White</td>
<td>Negro</td>
<td>American Indian</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Hawaiian,</td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>White</td>
<td>Negro or Black</td>
<td>Indian (American)</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Hawaiian,</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>White</td>
<td>Black or Negro</td>
<td>Indian (American), Eskimo, Aleut</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Korean, Vietnamese, Asian Indian, Hawaiian, Guamanian,</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>White</td>
<td>Black or Negro</td>
<td>Indian (American), Eskimo, Aleut</td>
<td>Chinese</td>
<td>Japanese</td>
<td>Filipino, Hawaiian, Korean, Vietnamese, Asian Indian, Guamanian, other Asian or Pacific Islander</td>
<td></td>
</tr>
</tbody>
</table>

*In 1850 and 1860, free persons were enumerated on the form for “free inhabitants”; slaves were enumerated on the form designated for “slave inhabitants.” For the free schedule, the instructions told the enumerators: “In all cases where the person is white leave the space blank in the column marked ‘Color.’” For the slave schedule, the listed categories were black (B) or mulatto (M).*

*Although this category was not listed on the census form, the instruction read:

5. Indians—Indians not taxed are not to be enumerated. The families of Indians who have renounced tribal rule, and who under State or Territorial laws exercise the rights of citizens, are to be enumerated. In all such cases write “Ind.” opposite their names, in column 6, under heading “Color.”

9. Color—Under heading 6, entitled “Color.” In all cases where the person is white leave the space blank; in all cases where the person is black without admixture insert the letter “B”; if a mulatto, or of mixed blood, write “M;” if an Indian, write “Ind.” It is very desirable to have these directions carefully observed.*

*In 1930, the census questionnaires included “Mexican” as a race category.

*In 1970, on questionnaires used in Alaska, the categories “Aleut” and “Eskimo” were substituted for “Hawaiian” and “Korean.”*
confusion generated among respondents to methods of racial and ethnic classification, especially among some Hispanic respondents (31). Amaro & Zambrana (1) present a thoughtful discussion of the complexity of both the collection and tabulation of multiple race/ethnicity data with Hispanic populations. As they note, the Hispanic population is extremely heterogeneous with regard to race, including individuals mixed with African, Native American, and Asian origins. Cautions are often given to health researchers that in national studies, data collected on the “Hispanic” population may, in fact, reflect information obtained from subgroups whose lives, resources, and health exposures are quite different from those of other Hispanic subgroups. This diversity among Hispanics also raises uncertainty as to how individuals of Hispanic background respond to questions of race and ethnicity. In the 1990 census, Hispanics were one of the largest groups to mark “other” for their race, suggesting that for some their notion of race was their ethnicity.

In response to these problems observed in the 1990 Census, the new standard established by OMB Federal Standards for Racial and Ethnic Data (91), contained three significant changes. First, there are now five minimum categories of race with Native Hawaiians and other Pacific Islanders separated out from the Asian and Pacific Islander category (45, 119). The new race categories are American Indian/Alaska Native, Asian, Black/African American, Native Hawaiian and Other Pacific Islander, and White. Second, and most important, individuals are now allowed to identify themselves as belonging to more than one racial group. Third, the placement of the Hispanic origin question was changed so that it now precedes the race question in order to increase the number of Hispanics selecting a specified race category (69).

Census 2000 was significantly affected by the new directive, especially the change in allowing individuals to “mark one or more” race categories on their forms (45, 70, 119). In the past, individuals of mixed racial heritage were forced to select only one race, no matter how they self-identified. Many of the advocates representing the multiracial community preferred this new explicit classification of each of the multiracial person’s races, rather than defaulting to a residual aggregate category of “multiple race.” This change, however, does not come without its challenges.

Implications of New Measurement Methods: The Example of Multiracial Classification

Tabulation and Bridging Tabulating multiracial respondents in the Census provides an opportunity to raise the counts of rarer populations, but it also raises the question of whether this occurs at the expense of misstating disparities in the health of specific racial/ethnic groups. This might arise often from pragmatic choices by planners or federal and state officials to reclassify multiracial respondents back into single, larger race categories. There are a number of reasons why this choice might be made including increasing statistical power and facilitating easier reporting approaches. But, by and large, a prominent reason might be for the purpose of bridging the data to other datasets where a single, larger race classification has been used (70, 75).
Reclassification into single-race categories may also for some racial groups create more meaningful classifications as compared to others. For example, some biracial individuals may strongly identify with one race over the other, depending upon their socialization or political consciousness about racial/ethnic identification (48, 49) or because of how they are treated in society based on their racial or ethnic status. Depending on which method is used for assigning their race, it may or may not coincide with the particular race most salient to the behavior of that individual.

Data from the National Center for Health Statistics’ (NCHS) National Health Interview Survey (NHIS) shed some light on how this works. NCHS predates the Census 2000 efforts in multiple-race data collection by over 15 years. Since 1982, the NCHS’ NHIS has been collecting data on multiracial persons in the United States. The NHIS questionnaire also includes a follow-up question asking multiracial persons which one race best represents their race. Results indicate that many multiracial individuals can identify with one race—particularly within American Indians/Whites, where most identified (75%) with being American Indian in the NHIS 1998 (75) as opposed to other races.

Information from another large survey, the California Health Interview Survey, a statewide survey of 55,000 individuals fielded for the first time in 2001 (98), takes the “most identify” question from NHIS a step further. The NHIS “most identify” question was asked only among those who had multiple responses for the question on the Office of Management and Budget (OMB) broad race categories (White, Black or African American, Asian, American Indian/Alaska Native (AI/AN), and Native Hawaiian and other Pacific Islander (NHOPI). In contrast, CHIS 2001 probed on the distinction of the Latino/Hispanic ethnicity from race, a federally mandated tradition since the 1977 OMB Directive 15. This action occurred, in part, because California’s Department of Finance’s (DOF) classification of race/ethnicity deems Latinos/Hispanics as a racial category mutually exclusive of other, more traditional racial categories, unlike the federal approach. The DOF provides the intercensal estimates of the state’s population, thus, to date, most state data reporting is by five race/ethnicity categories: White, Latino, Asian and Pacific Islander, Black, and American Indian/Alaska Native (98).

The methods used by the DOF involve a first tabulation of the Latino/Hispanic category. Federal OMB race categories are then assigned to the remaining non-Latino population, so that, for example, if one is Latino and American Indian/Alaska Native, that person is categorized as Latino. The result of this approach is inflation of the Latino/Hispanic count and fewer counts for the other race categories. While the effect is minimal among groups who are predominantly not Latino (e.g., Blacks or African Americans residing in California), the impact is considerable for California’s American Indians, since a sizeable number have dual Spanish heritage. This California DOF tabulation prioritizing Latino/Hispanic ancestry also affects Filipinos who report both Asian and Spanish ancestry.

Thus, even within a single-race data collection framework, tabulation rules can affect counts of a population. While counts are paramount in redistricting of legislative areas and in signaling growing political constituencies, in health data,
championing counts may be at odds with predicting the health status, risks, and health needs of some populations by race and ethnicity (70). For example, in measuring the uninsured rates among Latinos and non-Latinos in the United States, the question arises as to whether or not to count all persons who say they are of Latino or Hispanic origin in the denominator of the uninsured Latino estimate. Alternatively, excluding Latino persons who fundamentally view themselves as “White” from the denominator reduces the population size and, in this instance, is more likely to inflate the rate of uninsureds among Latinos. Another classification ambiguity arises in regard to persons who indicate being of American Indian/Alaska Native ancestry even though they are far removed from any connection with a tribal heritage. These complexities in race and ethnicity classification are clear indicators that how we measure and that the choices of the classifier can matter.

To further illustrate the implications of various approaches to race/ethnicity classification, we present simple, unadjusted tabulations of selected health access and health status indicators from the adult sample of CHIS 2001 crossed with four different methods of race/ethnicity classification definitions. In each of the methods, the race/ethnicity of individuals was classified into mutually exclusive categories of White, Latino, Asian, Black or African American, American Indian/Alaska Native (AIAN), Native Hawaiian and Other Pacific Islander (NHOPI), and other race. The four methods were: (i) Most Identify/Rarest, (ii) Rarest, (iii) Department of Finance method, and (iv) and adjustment of the DOF method using a Rarest adjustment. In the first classification method, “Most Identify/Rarest,” the race/ethnicity classification of multiracial and Latino respondents was based on their response to the “most-identify” question. If these respondents refused, said “both,” “neither,” or “don’t know” to the most identify question, they were then assigned to the “Rarest group.” In California, the hierarchy of rarest to most common racial/ethnic groups is: Native Hawaiian/Other Pacific Islander, American Indian/Alaska Native, Black or African American, Asian, Latino, and then White. By the “most identify/rarest method” a person reporting both Latino and White background who weighted these equally would be assigned to the Latino group. This method’s combination allocation rule foremost respects the choice of the respondent, but as a default, also elevates the counts of the smaller populations to some extent. The default “rarest” group decision is arbitrary and has greater statistical consequence than a “most dominant” group rule. But the method is sometimes used in minority health research because it increases the empirical chance of discovering the health needs of groups who are typically underrepresented in health data—especially American Indian/Alaska Natives, Native Hawaiian and Other Pacific Islanders, and Asian subgroups. Statistically this choice provides some opportunity to gain precision in estimates for rare groups about whom less is known in the field of public health.

The second method is classification by the “Rarest Group” decision criterion alone. In this instance, for example, an individual who reports both Black or African American and Latino background is assigned to the African American group. This approach inflates the sizes of rarest groups but does not take into account a
multiracial/ethnic individual’s sentiment in regard to racial/ethnic identification. This failure to consider individual identification is important in that emerging data indicate that self-reported ethnic identification is related to health outcomes (48, 49). This effect may be a function of greater precision that such information affords in estimating the sociocultural components of health, including health behaviors, diet, and neighborhood locations.

The third method conforms to the California DOF decision criterion discussed earlier. Here Latino ethnicity takes precedence over race, so that the other race/ethnic categories are defined as non-Latino White, non-Latino Asian, or non-Latino Black, for example. An individual who reports both Asian and Latino or Hispanic background, as many persons of Filipino ancestry do, is classified as Latino or Hispanic. In states such as California, the effect is to interject greater heterogeneity into the already diverse Latino/Hispanic population and to lose information about rarer subgroups, such as Filipinos and Native Americans. This heterogeneity of the Latino/Hispanic population, some argue, makes it difficult if not impossible to determine with accuracy the relationships among race, ethnicity, and health outcomes (1). With the sustained growth in the Hispanic population in the United States that is projected, continuation of this method of classification is troubling to some.

Finally, the fourth method is a variant of the DOF criterion, where all respondents who report being Latino or Hispanic are classified as Latino, except for American Indians/Alaska Natives, Native Hawaiians, and Other Pacific Islanders. These are the rarest racial groups in California. In this instance, the person is allocated to the rarest subgroup, no matter what his or her background with regard to Latino or Hispanic heritage.

The effect of these four methods of classification on estimation of race/ethnicity-associated health disparities varies across the different types of health outcomes assessed (see Figures 1–6). In addition, the greatest impact is generally observed
Figure 2 Percent of adults reporting no usual source of health care at time of interview by ethnic/racial background classifications (14).

Figure 3 Percent of adults reporting ever being diagnosed with asthma by ethnic/racial background classifications (14).

among the rarest groups. Also, because some of the variant classification methods hinge on the separability of the Latino/Hispanic origin/ancestry for some measures, the Latino population is affected as well. In the case of rare groups, given that these are precisely the racial/ethnic groups about which public health knows the least, the cost of imprecision in estimates could have serious repercussions on public health planning.

Increasingly, race/ethnicity reporting in public health needs to explicitly state the definition rules of race/ethnicity, particularly in states such as California where multiracial/multiethnic individuals comprise up to 4.7% (13) of the population.
Figure 4  Percent of adults reporting ever being diagnosed with a diabetic condition by ethnic/racial background classifications (14).

Figure 5  Percent of women, age 18 years and older, reporting a pap test in past year by ethnic/racial background classifications (14).

Figure 6  Percent of men, age 50 years and older, reporting a prostate-specific antigen (PSA) test in past year by ethnic/racial background classifications (14).
The acuteness of this problem is quickly becoming apparent as states struggle to be consistent with federal standards of racial and ethnic data and their own, only to find that the OMB guidance on race and ethnicity can misclassify their populations. Laws & Heckscher (68) in their study of racial and ethnic classification in six New England states found not only a range of ways by which individuals were classified but found several inconsistencies. For example, in Connecticut’s death registry, Creole was included under White, although in the South the person would be classified as having both African and White descent and if viewed as a Mulatto, classified as Black. Findings from the six-state study support the need to expand the OMB definition beyond the Hispanic/Latino definition to better classify persons who in the current system are misclassified (39, 68). We have offered here only four algorithms as an illustration, but other tabulations have been used elsewhere, such as assigning multiracial groups to the most dominant group. Lee et al. (70), Lucas (75), Parker & Makuc (97), OMB (92), and others describe various methods with similar outcomes and concerns. These methods vary in how a person’s multiracial response is assigned (e.g., to a single category termed whole assignment or to multiple categories termed fractional assignment) and whether or not assignment is based on deterministic/fixed rules or by some type of probability method (69, 75, 82, 92). OMB and the Census engaged in extensive discussions and sought public comment before arriving at a set of guidelines [see also Lucas (75), Lee (69), and Mays et al. (82) for fuller discussion of OMB 2000 tabulation methods] that can be found at: http://www.whitehouse.gov/OMB/inforeg/index.html#SP.

In a recent study designed to examine race-specific estimates of employer-sponsored health insurance utilization using different race allocation methods, estimates were similar across the groups with the exception of the AI/AN population (97). This study also illustrates, however, that demographic and socioeconomic differences between single- and multiple-race groups may affect estimates calculated using various bridge methods (75, 96, 97). The larger the differences in the sociodemographics of the two groups, the more likely that the choice of the bridge method matters. Underlying all of these decision rules are assumptions about the nature of race and ethnicity. The lack of an agreed-to standard also opens the door to classification choices that meet the researcher’s or agency’s needs or expectations but may harm comparability of findings across settings.

WHAT CAN WE KNOW ABOUT MIXED-RACE POPULATIONS?

Among some advocates and social scientists there is an increasing interest in the health concerns of individuals with a multiracial heritage (33, 42, 45, 103, 104, 114, 122, 133). At present, what we know about multiple or mixed race individuals is primarily limited to descriptions of the sociodemographics of the population (e.g., where they live, their level of education or income, which race/ethnicity pairings are most likely, rates of fertility, cohabitation, and marital dissolution (34, 48, 49, 59, 99, 100, 133).
But there are a number of methodological challenges in the use of multiple-race data. These include, most prominently, the difficulties in defining and sampling from an extremely diverse population. For example, self-identification as “multiple race” is vulnerable to the fluidity of racial identity, the principles of which have been the subject of social science research (43, 93, 103, 104). Situational ethnicity, a term coined by social scientists, refers to changing of race or ethnic identity usually within specific contexts. Racial identity is not necessarily stable and may be influenced by specific contexts (70). Factors that have been known to influence identity are where people live (the racial/density), their developmental stage of life, the context in which the question is asked, and perceived benefit or loss (122). This was apparent during debates among community groups as to the redistricting merits of choosing one or more races in completing the 2000 Census forms.

Despite the recency of work in this area, there may be important demographic distinctions between single- and multiple-race individuals in areas that are typically associated with differences in health status. For example, Corrin & Cook (25) found interesting differences between mono- and multiracial Black and White adolescents, depending upon their primary race identification. In looking at neighborhood differences, monoracial Blacks lived in neighborhoods with the highest level of poverty, followed by multiracial Blacks, multiracial Whites, and those who were monoracial White. Whether race-related risk indicators are additive or synergistic for multiple-race individuals is an open question (75). The level of complexity introduced in studies of multiple-race effects is clearly quite daunting but may offer opportunities to understand how it is that race comes to be associated with health.

SHOULD WE STILL TRY TO MEASURE RACE AND ETHNICITY?

When nations become more racially diverse, a natural evolution can occur in the measurement of race and ethnicity. For example, in 1976, the federal government mandated the inclusion by federal data collection agencies of Hispanic origin as an ethnic “overlay” to race (135). This mandate reflected the new immigration patterns that resulted in proportionally greater prevalence of Hispanic backgrounds and the emergence of Latino political power in the United States. In some instances, Spanish-origin groups in U.S. classification have been treated as equivalent to a racial group (as indicated above in the use of the DOF classification process), which raises statistical challenges that require careful evaluation.

Since 1992, the National Health Interview Survey, in response to burgeoning numbers of Americans of Asian/Pacific Islander heritage, expanded the number of Asian and Pacific Islander subgroup categories in its data files to accommodate cultural and linguistic heterogeneity with the aggregate AAPI category. In 1997, the Office of Management and Budget’s New Federal Standards for Race and Ethnicity Reporting mandated the separation of Native Hawaiian and Other Pacific Islander
from Asian as a race category. Most recently, as noted above, the Census 2000 allowed multiracial individuals to mark all of the races that apply to their racial makeup. Change in reporting standards invariably creates the opportunity for initial confusion and generates implementation costs. However, reporting developments since the 1964 Civil Rights Act are indicative of both political and societal will to preserve race and ethnicity reporting while at the same time adjusting classification methods to reflect temporal changes in how these constructs are construed at both the individual (e.g., self-labeling) and aggregate levels.

Lately, a somewhat provocative proposition on race is surfacing that suggests that a state government should discontinue its role in collecting race and ethnicity information (http://www.informedcalifornia.org; http://www.nlg.org/sf/rpi.html; http://dpls.dacc.wisc.edu/pubs/newsletters/sep02news.html#race; http://www.cpec.ca.gov/commission/Agenda0206/Tab_07.pdf (see page 8); http://www.adversity.net/RPI/rpi_mainframe.htm). Framed as a matter of privacy, the underlying premise is that government policies, in theory and in practice, have been successful in removing race-based barriers to higher education, employment, contracting, and certain human capital investments. The proposition’s current equal opportunity presumption suggests that there is little to be gained by continuing to measure race and ethnicity, and not enough to justify the privacy intrusion. This color-blind strategy has actually already been tried and abandoned many years ago by the state of New Jersey, which for a two-year period did not list race on vital certificates (35).

The perspective that government should drop its collection of race and ethnicity information from individuals comes at a time when public health is, itself, struggling with new complexities in the measurement of race and ethnicity. Perhaps public health might entertain the possibility of foregoing race/ethnicity classification difficulties and instead focus on alternative causes of disease, such as social class or income, to explain the race/ethnicity-associated health disparities that have been well documented. Public health might also strive to design effective health services for a diverse population without using race or ethnicity to guide its efforts. However, there are numerous examples from the empirical literature and experiences documented by public health agencies that suggest this is not a prudent approach.

One test of the need for race and ethnicity information in public health measurement is whether race differences are significant after controlling for indicators of social class, including income, education, occupation, and other socioeconomic, demographic, and geographic variables. The Institute of Medicine’s (IOM) “Report on Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care” compiled and documented numerous studies on health care access, use, disease patterns, and exposure to environmental hazards that show the independent effect of race as a significant explanatory predictor. The recently released materials from the IOM meeting on guidance in a national health care report on disparities also strongly supports that there is an independent effect of race (113).

Further, there is substantial evidence that differences between races exist in types of treatment, quality of care, and availability of care, even after controlling
for other factors. Several studies have demonstrated utilization and treatment variations by race among the Medicaid population, a means-tested public insurance program for low-income families (26, 28, 73, 106, 115, 124). Among patients with comparable private medical insurance coverage, there is evidence that racial/ethnic minorities receive inferior medical care (4, 16, 36, 47, 50). Thus, even within health insurance/income strata, race/ethnicity-based health disparities exist in access to health care. Further, there is evidence that harmful environmental exposures, factors that are clearly in the domain of public health, are to some extent associated with race/ethnicity. As an example, even after accounting for socioeconomic factors, race or ethnicity has been identified as an important reason for disparities in environmental harm (44, 46). Studies also point to the strong association between areas with high-level emissions of hazardous materials from industrial sites or waste management facilities and communities that have a high percentage of minority residents (85, 117).

Omitting race and ethnicity in public health data would erode the power of predictive models in public health. Also, omission of race/ethnicity information may introduce bias in estimations of the predictive power of other variables due to uncontrolled confounding. Income’s effect on health factors may be overstated or education’s effect understated. The end result would be missed opportunities in tailoring interventions to racial and ethnic groups with the most health risks. In the long run, scientific advances would be lost in understanding the role of race and ethnicity as both a risk and protective indicator in disease patterns, responses to treatment, and culturally based behaviors.

CONSIDERING RACE AND ETHNICITY IN CONTEXT: AN EXAMPLE OF CULTURAL COMPETENCE

Recently, there has also been much discussion of the issue of cultural competence in health and public health interventions (9, 18, 27, 41, 71, 83, 108). Cultural competence, in this instance, refers to incorporating expertise, some drawn from research findings on race and ethnicity, into optimal public health practice. We offer an example of a cultural competence approach below, in part to demonstrate what would be lost should public health entertain notions of foregoing the measurement of race and ethnicity.

At an individual level, the literature on racial/ethnic variations in health and health outcomes suggests three major considerations in applying knowledge of an individual’s race/ethnicity in the healthcare setting. First, race/ethnicity, within the context of knowledge of other patient characteristics, may be used to guide the medical and social history-taking for disease-specific risk factors with known racial/ethnic group associations. For example, rather than assuming all Asian and Pacific Islanders require tuberculosis screening, a practitioner might use a social history of immigration from a country with a high tuberculosis prevalence (a more proximate and stronger risk indicator for tuberculosis infection) to guide screening
decisions. Additionally, rather than approaching all African Americans as being at high risk for sexually transmitted infections (STI), practitioners might incorporate questions about STI risk behavior into their history-taking for all patients and consider the base rates of STIs in the clinic setting. This approach requires that practitioners educate themselves about disease mechanisms and risk factors for which race/ethnicity serves as a proxy.

Though much more research is still needed in this area (67), there is currently sufficient knowledge for individual’s race/ethnicity to partially inform preventive health screening recommendations. For example, given the significantly higher prevalence and earlier incidence of prostate cancer in African Americans as compared to Whites, the American Cancer Society recommends initiating prostate cancer screening in most men at age 50, but in African American men at age 45 (2). Translating into practice this inability to more clearly delineate which men are at risk for prostate cancer means that (at least for prostate cancer screening) a 45-year-old African American male should be treated differently from a similar-age White male. However, race/ethnicity alone, similar to other patient characteristics, is not informative enough to characterize the specific risk of a given individual from that racial/ethnic group.

A culturally sensitive approach goes beyond consideration of race/ethnicity main effects in health disparities and requires eliciting information about an individual’s values, health beliefs and practices, communication preferences, and preferences for interacting with the healthcare system, among other factors. As with using an individual’s race/ethnicity to tailor medical history-taking, this information is used to inform but not to dictate a specific approach.

A culturally sensitive approach also recognizes that practitioners live within their own cultures and could profit from exploring their own potential biases in approaching individuals from different racial/ethnic groups. In 2002, the IOM reported on the differences in the kinds and quality of health care received by racial and ethnic minorities and nonminorities in the United States (110). The IOM report found that individual differences among health care providers in the effects of patients’ race or ethnicity on their clinical judgments contribute to health care disparities (110). Many of these differences reflect racial or ethnic stereotypes. Independent of medical necessity, a patient’s race/ethnicity has been shown to influence doctor-patient communication, recommendations for cardiac catheterization, and intensity of hospital services (23, 107, 134). As Bloche (2001) notes, the inherent uncertainties of diagnosis and treatment provide fertile ground for judgments that are influenced by personal biases as well as medically informed opinions.

Individual approaches to health care decision-making have corollaries at the population level where the field of public health typically has its greatest involvement. Knowledge of the racial/ethnic distribution of the population of interest, within the context of knowledge of other characteristics of the population, may be used to guide estimates of disease burden, needs for services, and optimal methods of intervention. The benefit at the institutional level of collecting race/ethnicity
data (within the context of collecting data on other population characteristics that contribute to health as well) is to reduce uncertainty about health needs and the anticipated burden on the health care system. Such knowledge can also assist in assuring the availability of resources for culturally appropriate interventions. At a basic level, this translates into providing culturally appropriate patient education literature in the opportune location at the optimal time. Institutions can also use racial/ethnic information for quality assurance purposes to monitor for systematic patterns of care that are indicative of nonmedically warranted choices at the individual level. As an example, if an institution collects case mix–adjusted race-specific data on the proportion of patients receiving cardiac catheterization, then it can use this information to target variations in processes-of-care that should be standardized to improve quality of care for all patients. This is a pragmatic approach to eradicate racial disparities in health care delivery that systematically disadvantage racial/ethnic minorities (8). Methods of achieving change include rule-based cost control mechanisms, modification of financial incentives for physicians, and strengthening of doctor-patient relationships.

In general, public health interventions may be more effective when targeted to specific populations in a culturally appropriate manner (52). Interventions that have been successfully tailored to specific populations include: improving diabetes education for Mexican Americans using community health workers (11), improving blood pressure control for a low-literacy hypertensive population using a worksite tailored educational program (37), increasing fruit and vegetable consumption among rural African American church members (15), and increasing early detection of cervical cancer among Native American women (84). A mechanism common to many community-based interventions is involving the community in identifying its needs (111) and training community members to serve as lay health workers, or consejeras, to deliver the intervention (7, 24, 87, 111). Knowledge of the risk profile and resources of a community also allows for identification of public health interventions that target specific barriers to health within that community. For example, lack of availability of healthy foods for people with diabetes could inform nutrition interventions for such communities. These interventions might range from educating patients within these communities about alternate sources for healthy foods, to increasing food distributor and storeowner responsiveness to dietary needs of community residents.

Finally, methods of cultural competence in public health practice include recognition that geographically expressed racial/ethnic segregation patterns also have implications for creating health disparities that are linked not necessarily to the race/ethnicity status of the individual but rather to the community. For example, several recent studies directed toward assessing the independent contribution of residential racial segregation suggest that, independent of the race/ethnicity of an individual, there is a health risk effect associated with the racial distribution of the community (22, 54, 128). Using data from the National Longitudinal Mortality Study, Jackson and his colleagues (54) found that, after adjusting for family income, age-adjusted mortality risk increased with increasing minority residential
segregation among Blacks aged 25 to 44 years and non-Blacks aged 45 to 64 years (54). Further, the availability of appropriate foods for people with diabetes (e.g., low-fat milk, high-fiber and/or low-carbohydrate bread, fresh fruit, and green vegetables) may be lower in inner-city neighborhoods. In a recent study (51) of grocery stores in East Harlem (a multiethnic lower-income neighborhood in New York City) and in the Upper East Side (the predominantly White, relatively affluent adjacent neighborhood), only 18% of stores in East Harlem, in contrast to 58% on the Upper East Side, had at least one of the diabetes-healthy food items on their shelves. This limited availability of diabetes-healthy food items within East Harlem likely affects people with diabetes from all racial/ethnic groups who reside within that neighborhood.

LOOKING TO THE FUTURE

The diversity of the United States with growing numbers of interracial unions, immigration, and changes in the way that individuals view their racial identity and ethnic heritage guarantees that the problems of classification of the population will get more complex (31, 48, 49). This complexity comes at a time when there is a greater demand for public health to better serve its communities through targeted interventions to reduce morbidity and mortality. The one-size-fits-all health promotion, health campaigns, health services, or even statistical approaches to health data are no longer acceptable to racial and ethnic minority groups. Although there are no ready answers to the issues that confront the field, a number of approaches and strategies might be useful to contemplate. First, in light of the growing racial and ethnic diversity in the United States and the desire from many agencies for uniformity across datasets, it is critical that careful thought be given to the range of racial and ethnic backgrounds that occur, including admixtures of these entities. Some would go as far as to recommend that ethnicity, as opposed to race, serve as the basis for classification (68), but this solution must be evaluated empirically prior to widespread adoption if it appears warranted. Multiracial populations will become increasing larger, behooving us to develop a better understanding of the ways in which multiracial statuses are associated with health and better methods of preserving respondent confidentiality when reporting information from very small groups. Also, whichever categories are developed need to suitably capture foreign-born immigrants and their offspring because we know that health statuses and risk may differ as a function of foreign-born status. Research is also needed that assists us in understanding from a public health perspective the factors, including intrapsychic, interpersonal, and situational, that affect race reporting both from the participant as well by the health system where much race and ethnic reporting occurs. We need to clarify the nature of race and ethnicity as a risk indicator of health, as well as its influences in the context of other correlated factors such as geography, urban density, immigration status, age, gender, and social class. Further, given the emerging body of research on
MA YS ET AL.

ethnic/racial health disparities, now is the time for us to ask new questions, including how racial and ethnic statistics are being used in developing health policies, in the delivery of health care services, and in the promotion of public health? Are there better ways in which we can use what we know to reduce racial and ethnic health disparities in this country? Answers to these questions will not come easily. There is no one method of allocation that will be best for either a population or a data user, but rather there are advantages and disadvantages associated with each method. Federal guidance that can bring consistency of methods used, and clarity of the advantages and disadvantages of choices is critical to moving this agenda ahead. Researchers and planners need to be aware of these complexities, particularly when using bridging methods across multiple data sources.

RECOMMENDATIONS

While in many regards Public Health’s emerging understanding of the complexities of measuring race and ethnicity in relation to health is by no means complete, at this point we respectfully suggest the following recommendations:

1. Leadership in the collection of race and ethnicity data by the federal government is critical. This leadership will affect both the quality of the census as well as a variety of other datasets that support and contribute to effective public health programs. The reach of this leadership will be broad ranging from the development of population denominators to accurate administrative hospitalization information. We suggest:
   A. Whenever possible, self-report of race and ethnicity is the preferred method for data collection (39, 68, 69).
   B. Whenever feasible, allocation methods should collect and respect the self-identification of multiracial individuals to guide tabulation decisions into single-race categories (69).
   C. Attention must be paid to allocation methods as some may disadvantage populations (particularly American Indians/Alaska Natives) or advantage other populations (69, 82).
   D. For data users, researchers, state and local health departments, and collectors of federal data (i.e., funeral directors, hospitals, nursing homes, Bureau of the Census), research, training, and policy agenda must be facilitated on the implications of the choices of various bridging and allocation methods for particular racial and ethnic populations to enhance the quality of the data collected and to develop consistent, reliable, and valid policies across and within federal data activities (39).
   E. A scientific agenda must be developed that addresses the fluidity of racial and ethnic identification, and the relationship to data collection, provides an evidence base for clinical practices, provides direction on how to interpret health estimates for multiple race groups, and provides
a better understanding of health estimates, health outcomes, health processes, and access to care for single-race versus multiple-race persons (19, 70, 74, 82).

2. Recognizing its major role in ensuring the public health, the federal government should develop a universal taxonomy of race and ethnicity prior to the collection of the decennial census. This will support efforts at both the national and international levels to achieve appropriate data comparisons bridging multiple data sources. We suggest that this might require a classification system

A. that is characterized by at least 3 properties: 1) consistent classification principles; 2) mutually exclusive properties; and 3) flexibility to absorb entries not yet identified (10, 38, 69, 70).

B. that addresses and brings clarity to use of race and ethnicity in a manner that is more consistent with how individuals think of themselves. This approach recognizes that these categories are most likely as much a function of political and social policies both in the United States and abroad as they are a function of biologically determined factors.

C. that moves the field of public health beyond its current conceptual and methodological framework in regard to race. At present, race is often seen as a biologic indicator whereby a specific group of people who share a genetic heritage are genetically different from other members of other racial groups. Though race is a risk indicator useful in predicting health status, outcomes, and treatment responses, it is unclear to what extent this is a function of biological factors versus cultural, social, economic, or geographically based environmental factors (38, 60, 127, 130).

D. that results in availability of denominators, congruence of denominators with numerators, facilitates consistency of response by participants, and consistency in methods for the use of the data (6, 82).

3. Although measurement of race and ethnicity has a long history in the field of public health, it is time to move beyond reductionist biologic or genetic perspectives. We have an exciting opportunity to explore relationships between racial and ethnic populations and their social and physical environments. Further, we soon will have the opportunity to enlarge our focus to their interactions with genetic influences in predicting health outcomes. To take full advantage of these opportunities, we need in our research and practice efforts to

A. include data collection on complex social variables for which race or ethnicity is often used as a proxy. This might include social status, neighborhood context, perceived discrimination, social cohesion, social capital, social support, types of occupation, employment, emotional well-being, and perceived life opportunities (10, 29, 30, 62, 63, 65, 66, 77, 78, 79, 81, 130). New methods of measurement, such as geocoding, may prove especially helpful for some of these areas. The fuller assessment of potential risk indicators may serve to clarify the direct effects of race and
ethnicity on health. At the same time, it may allow development of alternative frameworks of conceptualizing health disparities and ultimately new directions for public health interventions.

B. measure the construct of ethnicity more broadly than simply Hispanic or Latino (1, 12, 39, 68, 129).

ACKNOWLEDGMENTS

This work was supported in part by the National Institute of Mental Health (MH 61774) and the National Institute of Drug Abuse (DA 015539). We would like to thank Drs. Dale Hitchcock and Olivia Pokras-Carter for their comments. We take all responsibility for the information in the manuscript.

The Annual Review of Public Health is online at http://publhealth.annualreviews.org

LITERATURE CITED


113. Swift EK. 2002. Guidance for the
110  MAYS ET AL.

114. Tafoya SM. 2000. Check one or more...mixed race and ethnicity in California. Calif. Counts 1:1–11
ing and Improving Health, 2nd ed. Washington, DC: US GPO
fills and Their Correlation with Racial and Economic Status of Surrounding Commu-
119. Wallman KK, Evinger S, Schechter S. 2000. Measuring our nation’s diversity: develop-
ing a common language for data on race/ethnicity. Am. J. Public Health 90(11):1704–8
132. Wood AJJ. 2001. Racial differences in the response to drugs—pointers to genetic dif-