The concept and measurement of race and their relationship to public health: a review focused on Brazil and the United States

O conceito e mensuração de raça em relação à saúde pública no Brasil e nos Estados Unidos

Claudia Travassos [1]
David R. Williams [2]

[1] *Departamento de Informações em Saúde, Centro de Informação Científica e Tecnológica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*
[2] *Department of Sociology and Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, U.S.A.*

Correspondence
Claudia Travassos
Av. Brasil 4365,
Rio de Janeiro, RJ
21045-900, Brasil.
claudia@cict.fiocruz.br

Abstract

Race has been widely used in studies on health and healthcare inequalities, especially in the United States. Validity and reliability problems with race measurement are of concern in public health. This article reviews the literature on the concept and measurement of race and compares how the findings apply to the United States and Brazil. We discuss in detail the data quality issues related to the measurement of race and the problems raised by measuring race in multiracial societies like Brazil. We discuss how these issues and problems apply to public health and make recommendations about the measurement of race in medical records and public health research.

Race; Medical Sociology; Equity; Review Literature; Skin Color

"*Race is a social construct, but as for other aspects of social stratification, with biological consequences*" [1].

The notion that health is influenced by the social position of individuals has been known for many centuries. Nancy Krieger [2] notes that since Hippocrates the relationship between health and social position has been acknowledged. It has also been shown that social disparities in mortality exist for almost all causes of death in most societies, and these disparities have been increasing in recent decades in several developed countries [3].

Race has been used extensively in the medical and public health literature, especially in the United States, to measure social differences in health outcomes and treatment, and its use has increased in recent decades. In the US, there is a vast literature that relates race to disparities in health outcomes, which shows that race is an important predictor of health status. "Blacks" in the US are disadvantaged compared to "Whites" on most indicators of economic status and health. Despite a reduction in these racial inequalities on both of these indicators during and immediately after the Civil Rights movement (the mid-1960s to the mid-1970s), they have remained large or have widened ever since [3]. In the US, adjustment for socio-economic status (SES) always reduces and sometimes even eliminates racial disparities in health. A recent publication of the Institute of Medicine

also documented that there are large racial differences in the quality and intensity of medical treatment in the US, even after adjustment for access factors, SES, and severity of illness [4].

In Brazil, there are fewer studies of racial inequalities in health. Batista [5], using data from death certificates, has shown that "Black" men and women had the highest crude mortality rates in 1999 in the State of São Paulo. Data based on census and national household surveys show that aggregate infant mortality in Brazil in the years 1977, 1987, and 1993 was higher for "Blacks and *Pardos*" ("Browns") and that it declined at a lower rate when compared with "Whites" [6]. Martins & Tanaka [7], using data from the Committee on Maternal Mortality, have also shown large differences in the risk of dying due to maternal causes in the State of Paraná in the years 1993 and 1997, which disproportionately affected "Black" and "Yellow" (Asian) women. Maternal mortality did not differ between "*Parda*" ("Brown") and "White" women. Dachs [8], using data from the 1998 National Household Survey (PNAD), found no statistically significant difference by "skin color/race" in self-assessed health status after adjusting for education and income level. Barros et al. [9], based on longitudinal data, have shown worse health outcomes for "Black" children in Southern Brazil, which is reduced after adjustment for SES and various other variables (marital status, maternal age, parity, planned pregnancy, social support, smoking, work during pregnancy, and antenatal care). The study results also suggest that "Black" mothers receive lower quality of care as compared to "White" ones. There are also indications that in Brazil racial inequalities are more common in treatment than in access to health care services [10].

The objective of this article is to review the literature related to the concept and measurement of race with a focus on the US and Brazil. We will discuss both the measurement of race in these two multiracial societies and data quality problems. We also make recommendations about the measurement of race in medical records and public health research. Although the use of race in public health research has been discussed in relation to definitional and methodological problems in the United States, the Brazilian public health literature has not discussed in detail how such problems apply to Brazil. This article is intended to review the literature and introduce a discussion regarding broader as well as country-specific questions and problems related to the use of this category in public health.

The race concept

Numerous authors have argued recently that race is a social construct, with only limited (if any) biological meaning. Human race is viewed as a product of our history and culture and not as a marker of our genes. As J. Kaufman [11] (p. 101) argues, race is a product of our cultural imagination: "…*race is not in our heads because it is real, but rather it is real because it is in our heads*". Thomas LaVeist [12] reviewed the definitions of race in various medical dictionaries and showed a lack of scientific rigor in such definitions. David Williams [13] reviewed the changes over time in the definition of race in biomedical and social sciences dictionaries (anthropology, psychology, and sociology). In most but not all of the social science dictionaries, race is viewed primarily as a sociopolitical construct with strong cultural components. In contrast, all biomedical dictionaries except one defined race as reflecting genetic traits and thus ascribed a dominantly biological meaning to race [13].

In the early 19th century, the various fields of science had adopted the ideological public view of human differences based on race. This ideology was linked to colonialism and slavery, which (especially in the United States) established a mode of classification based on a rigid hierarchy of socially exclusive categories that divided Europeans, Africans, and Indians (http://www.aaanet.org/stmts/racepp.htm, accessed on 13/Aug/2002).

The concept of race began to be questioned in the late 1930s by Ashley Montagu, culminating with his book *Man's Most Dangerous Myth: the Fallacy of Race*, published in 1942, in which he claimed that race was a biological myth [1]. After World War II, Montagu wrote the final text of the UNESCO *Statement on Race* that denied any scientific meaning for the concept of race. Many physical anthropologists and biomedical scientists strongly opposed this document [14]. However, a study of the definitions of race in physical anthropology textbooks published between 1932 and 1979 clearly showed that the entire field moved towards Montagu's view over time [15]. By the 1970s the modal position of these texts was that human races do not exist. However, a few physical anthropologists still hold the traditional view. For example, in the

1985 edition of the *Anthropological Glossary* [16] race was defined as "*a genetically distinct inbreeding division within species*". However, at the end of the 20th century the Executive Board of the North-American Anthropological Association, based on recent scientific evidence, but without consensus from all members, officially stated: "*physical variations in the human species have no meaning except the social ones that humans put on them*" (http://www.aaanet.org/stmts/racepp.htm, accessed on 13/Aug/2002).

In the evolutionary field, race is assumed as a synonym for subspecies. Similar to race, subspecies is also an imprecise concept. The traditional definition relates subspecies to a geographically circumscribed and genetically differentiated population, but from an evolutionary genetic point of view the problem is "*that many traits and their underlying polymorphic genes show independent patterns of geographical variation*" [17] (p. 632). One argument, based on the traditional concept of subspecies, frequently present in the debate about the subdivision of the human population into "races", is that most of human genetic diversity exists as differences between individuals within populations. Amongst other studies, Rosenberg et al. [18] have shown that only 3 to 5% of the existing genetic diversity can be used to distinguish what could be called as human "races". The genetic diversity between groups of humans is the lower than that of several other mammalian species and is considered to be too small to allow a distinction of humans subspecies under the traditional concept [17].

Nonetheless, methods traditionally used (Fst and related statistics) to measure levels of human genetic diversity are not sufficient to discriminate contrasting theories about human evolution, because in modern evolutionary science, subspecies must be genetically differentiated due to barriers to genetic exchange that have persisted for long periods of time. Templeton [17] uses the concept of "genetic distance" to test contrasting evolutionary models. The Candelabra Models (old and new) imply that major human geographical populations (Africans, Asians, and Europeans) are branches of a candelabra and therefore valid "races", while the Trellis Model poses recurrent genetic interchange among Old World human populations in such a way that there was no separation into evolutionary lineages and as a result there is no such thing as human subspecies or races. Both models agree with the African origin of anatomically modern humans. Templeton presents evidence that genetic distance models do not fit treeness as required by the Candelabra Models,

but fits well to a restricted gene flow model (Trellis Model). Therefore, genetic distance does not validate the concept of human races as evolutionary distinct lineages. By using modern genetics, Templeton shows that human beings are not biologically distinct groups.

Human classifications based on physical traits are also considered not to have any evolutionary validity. They represent adaptive traits of human population to the environment. Physical traits such as skin pigmentation, hair color and texture, and the shape of one's nose or lips are erroneously used as markers of race. They are assumed to be adaptations to geographic factors such as solar radiation and temperature [19]. Melanesian and African populations share traits such as dark skin, hair texture, and cranial-facial morphology, used to classify people into "races", but they have great genetic diversity. Europeans are genetically closer to Melanesians and Africans than Africans are to Melanesians [17].

Contrary to the Mendelian tradition, which assumes that for any one gene there is only one phenotypic outcome, modern genetics shows that the phenotypic expression of any gene may vary expressively depending on the environment (the organism's genome as well as external factors). Mendelian ratios (three-to-one ratio) are the result of very special cases, in which the phenotypic expressions of enzyme pathways are little influenced by the environment. This is because they might express very trivial traits of the phenotype [20].

The complexity of the expression of most genes, which relates to the other genes in the organism's genome, the cellular and extra-cellular environment and the environment outside the organism [20] imply that the number of diseases that follow the Mendelian tradition is also limited. One example is sickle-cell anemia, but the majority of diseases do not represent a single-gene disorder, and molecular genetics is still far from being able to provide further understanding of chronic illnesses.

Despite the fact that race has been used as a surrogate for genetic information until the onset of molecular genetics, there is no scientific support to continue using race in Public Health as a marker for genetic susceptibility [21]. Parra et al. [19] have recently shown that skin color in Brazilians cannot be used as a genetic marker, because physical traits have been shown to be a poor predictor of African ancestry in this population. In both the United States [21] and Brazil, although the risk of sickle-cell anemia varies by race, race is not a reliable predictor of sickle-cell anemia.

However, the controversy in the biomedical literature in relation to race remains alive. Neil Risch [22] recently argued in favor of the validity of using racial/ethnic self-classification in biomedical research. Their arguments are based on the acceptance of an evolutionary tree for human races. Race is defined by these authors as the person's primary continent of origin based on the evolutionary tree, while ethnicity is viewed as a self-defined construct with geographical, social, cultural, and religious meanings. For these authors race is a biological concept, while ethnicity is a social construct that could be used as a proxy for race/genetic variation. In opposition to the conclusions of a recent study on drug response, which proposed that genetically inferred clusters are more informative than commonly used ethnic categories [23], Neil Risch and colleagues claim that ethnicity or ancestry are actually more genetically informative than clusters based on the analysis of random genetic markers. The argument in favor of ethnicity is that the number of *loci* required in genetic data to discriminate more recently separated populations (or those influenced by admixture or migration) is much greater than that needed to discriminate populations with ancient separation. However, these arguments must be interpreted in relation to the validity of the evolutionary theory used that assumes the existence of "races" in the human population. Secondly, as will be discussed later in this paper, there are validity and reliability problems with racial taxonomies. Methodological problems are likely to increase when the taxonomy is applied across countries and to admixed populations. These are major limitations given that biomedical research should generate information with the greatest possible validity across populations.

Another recent article [24] based on a small sample of southern Brazilian women suggests that the self-referred number of "Black" ascendants is a reliable marker of susceptibility to certain heritable health conditions.

Despite existing controversies in the biomedical literature, it is widely accepted that racial/ethnic categories are imprecise and changing measures that are historically, administratively, and politically constructed. The salience given to race, as well as the meaning and the measure of race itself in census and health data, varies across countries and across time. The history of race classification in the US and Brazil are good examples of these variations as will be discussed later in this article.

Race and class

The salience of race in measuring social inequalities varies across countries. The United States, a highly racialized society, has been measuring race since its first census in 1790 [25]. In this census, there was no information on social class (occupational class), but people were classified by race. Race has been more salient than class in the US, and in some ways this supported an ideology of a classless society [2].

In addition, the importance given to race in US health statistics is not universal. Studies of health disparities in European countries have attributed much more salience to the concept of social class than to race. In the US, until recently, age, sex and race were the routinely used variables in reports on vital statistics. In contrast, social class data have been linked to health information for more than a century in England, although there is growing attention to race with both the Census of England and Wales. In Portugal, information on race is not collected in the National Census.

Differences in demography and ideology are identified as important reasons behind the variables selected to be included in health data systems [2]. Ancestry and therefore race had limited influence among public health professionals in the early 20th century in Great Britain, and preventive medicine looked for historical and social explanation of health, in contrast with what happened in the US [26]. Only recently, in 1991, was a measure of ethnicity included in both the Health Survey of England and Wales census [25], given the extensive migration of people from former colonies in the second half of the 20th century, ethnic minority groups now constitute 6 percent of the population of England and Wales. As this population continues to grow, there is increasing interest in racial/ethnic variations in health in the UK, as is happening in other countries. In Brazil, only in 1996 did it become mandatory to include data on "race/skin color" on the death certificate and the Information System on Live Births (SINASC).

As for other areas (variables) of social life, race is a particular dimension of social stratification which defines differences in access to goods and services. Race may be related to social class but is different from it, even though both concepts carry socially constructed meanings. Race is based on the physical (color/race) characteristics of individuals, while social class is a product of social relations.

Marger [27] highlights that the unequal distribution of resources within a society creates a *system* of stratification, meaning that in mod-

ern societies stratification occurs along several dimensions, with class stratification as the most prominent. Multiethnic societies are also stratified in relation to ethnicity. However, as this author argues, "*the empirical relationship between class and ethnicity is complex and subject to frequent change*" [27] (p. 63). According to E. O. Wright [28], race represents non-class forms of oppression and can reflect relations that are quite independent of class. Candace Nelson and Martha Tienda [29] argue that although ethnicity cannot be reduced to a class phenomenon, social position plays an important role in molding ethnicity over time and place.

Socio-economic status (SES) is typically measured by income, education, and occupational status or some combination of these three measures [30]. Variations in health status across SES in the United States are in general larger than racial ones, but race tends to predict increased health risks independently of SES [25,31]. Race and SES are correlated but not identical. For example, "Blacks" are three times more likely to be poor than "Whites", but two-thirds of blacks are not poor and two thirds of all poor Americans are white [30]. Similarly, in Brazil, race and SES are not equivalent, and racial disparities are smaller than SES ones. Data from the 1998 PNAD show that income is unevenly distributed across race categories: "*Pardos*" are largely concentrated in the poorest quintile, followed by "Blacks". On the other hand, Asians ("Yellows") are largely concentrated in the richest quintile, followed by "Whites". In the Northeast, the poorest region in the country, "Whites" are largely concentrated amongst the poor. Gender inequalities in the labor market have also been shown to be greater than racial inequalities in Brazil [32].

Measures of race

The US experience

Race is central to the organization of life in the United States, which was the first country to collect census data on race. Racial categories in census data in the US have changed regularly, such that no racial classification scheme has been used in more than two censuses [25]. Prior to 1960, race was assessed by interviewer observation. Since then, the assessment of race has been self-reported. A new classification was used in the 2000 census. The main characteristic of the US race classification is that it is mainly based on ancestry and not on phenotype.

From the very beginning, racial categorization was an expression of social status (value) of particular groups in society. Race was considered by the "White" elite as a natural distinction in human identity [33]. Following the US Constitution, "Blacks" were counted as three-fifths of a person [25], and until 1850 "Blacks" in the US census were categorized as either "slave" or "free colored". Early censuses did not count Indians unless they were "Civilized" (the latter being those who did not live on reservations and who paid taxes). The 1870 census classified the indigenous population as "Pure Indians" and "Half-breeds" [34]. Only after 1924, when American Indians were given citizenship, they began to be classified in a single racial category according to the US census [25].

The 1850 and 1860 censuses used the categories "Black" and "Mulatto" (tabulations would aggregate under the term "colored") for free African descended people, and from the 1910 to the 1930 census the term "Mulatto" was used again together with "Negro" to classify African descendents [34]. The temporary usage of a term [mulatto] to classify admixed people in the US census supported the polygenist theory of the superiority of "Whites", which additionally contended that hybrid racial species were less fertile and had shorter life spans than pure-race persons [33]. For this purpose, the class "Mulatto" was defined as including anyone having any percentage of African blood. "Mulatto" was perceived by the color of the skin by census enumerators and was not based on genealogical history. It referred to people in whom the mixture of "White" and "Black" was visible [35].

The 1890 census "refined" this admixed racial category: besides "Mulattos" it included the categories "Quadroon" (one Black grandparent or one mulatto parent and the other white) and "Octoroon" (a Black great-grandparent or one Quadroon parent and the other White) to further distinguish the level of Black blood. The "Mulatto" category remained in the 1910 and 1920 censuses, but was dropped in 1930 by census officials who claimed it was inaccurate [33]. Consistent with racist laws, the terminology for admixed populations was eventually substituted by "Non-White" categories based on the "One Drop of non-White Blood" rule. This rule stated that a single Black ancestor would classify a person as Black, despite appearance. The 1930 US census stated that "*a person of mixed White and Negro blood should be returned* [classified] *a Negro, no matter how small the percentage of Negro blood*" [33] (p. 1741). The ideology behind the "One-Drop" rule was one against miscegenation [33] and shaped the "Black" and "White" racial divide of the US population. Many Southern states in the US

had laws that prohibited interracial ("Black-White") marriage until the US Supreme Court ruled such laws unconstitutional in 1967. It was not that admixture did not exist, but that it was seen as something to be prevented and reduced. From then until the 2000 census, racial classification in the United States did not give any room for multiracial classification. Nonetheless, in earlier censuses, a small number of persons checked the "other" race category and specified that they belonged to multiple-race categories.

Influenced by the 19th century notion of race in anthropology, race categories were changed in the 1900 census to match what were considered the four major races. Discrete categories were created to express mainly the continent of ancestry: "Caucasian" or "White", "Negro" or "Black", "Mongolian" or "Yellow", and "Indian" or "Red" [34]. As such, racial categorization gained scientific status.

"Whites" are the large majority of the US population. In the 2000 US census "Whites" corresponded to 75% of the US population. The distinctiveness of "Whites" as the dominant race is reflected in the US census categories in the fact that this group was always separated from "Non-Whites". For a short period of time in the 19th century the US census term "Colored" included all races except "Whites" [36]. Unlike the other racial categories, the term "Whites" never changed, although its meaning did. In 1910 a category "All Others" was created that included Mexicans, but in 1930 they became counted as "Whites".

The use of "Chinese" and "Japanese" in early racial classification in the US shows that the lack of distinction between the concept of race and ethnicity was present in the measurement of race in the US since its early times. "Chinese" was a separate racial category in the 1869 census. Later it was merged into the "Colored" group, but in 1900 it became a sub-category among the "Yellows" or "Mongolians", with the "Japanese" as another sub-group. Other Asians were counted but rarely presented in published tabulations [34].

The Civil Rights Act of 1964 and 1968 required racial/ethnic data for monitoring policy. In response to these needs, the US Office of Management and Budget (OMB) established a race classification standard in 1977 to be used by all the statistical agencies of the Federal government, including the US Census. These guidelines were used in the 1980 and 1990 censuses. The OMB classification officially introduced ethnicity in the US race classification. It contained a minimum set of four race categories ("North-American Indian or Alaskan Native", "Asian or Pacific Islander", "Black", and "White") and two ethnic categories ("Hispanic" and "Not Hispanic origin").

"Hispanic" was defined as a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. Race and ethnicity could be collected separately or in a combined format. In the separate format "Hispanics" were also to be classified as "Black" or "White", and in the combined format these two "races" were to exclude "Hispanics" (http://www.whitehouse.gov/omb/fedreg/ombdir15.html).

At the time of the establishment of OMB standards for racial classification, civil rights advocates accepted the definition of race as it was, arguing that since it was the basis for discrimination, it should be assessed to monitor the success of policies to reduce discrimination [33]. As a result, the dichotomy between "Blacks" and "Whites" based on the racist "One-Drop" rule and the absence of multiracial categories remained in the US official taxonomy.

In the 1990s the OMB reviewed its race standards and adopted a revised classification system in 1997 that was applied in the 2000 census. Under the new standards, as in earlier ones, the OMB does not make a clear conceptual distinction between race and ethnicity or completely rule out biology from them: "*racial and ethnic categories should not be interpreted as being primarily biological or genetic in reference. Race and ethnicity may be thought of in terms of social and cultural characteristics as well as ancestry*" (http://www.whitehouse.gov/omb/fedreg/ombdir15.html, accessed on 23/Aug/2002).

Various changes were introduced in the definition of the categories, in question format (the use of separate questions to inquire about race and ethnicity is recommended, with the ethnicity question coming first), and in terminology. The new classification provides a minimum standard with five race categories ("American Indian or Alaska Native"; "Asian"; "Black or African American"; "Native Hawaiian or other Pacific Islander"; and "White") and two ethnic categories ("Hispanic or Latino" and "Not Hispanic or Latino").

The question of how to measure people with multiracial ancestry was also discussed, but a "multiracial" category was not included. There was strong activism on both sides of the issue, but in the end the OMB decided to maintain race classification based on discrete race categories. As race classification reflects people's consciousness about race, but also acts as reifying race in people's consciousness, the

OMB taxonomy has played a role in perpetuating the idea of "pure" races in many people's minds. However, to deal with people of multiple racial heritage, respondents are now allowed to check as many responses as are applicable in response to the race question [37].

The concept of race based on ancestry in the US racial taxonomy and the absence of a multiracial category is likely to raise difficulties for many admixed people. In the census question about race, the respondent can specify when he or she selects "some other race". In the long form of the 2000 census questionnaire, applied to a sample of the US population, people also answer a question about their national origin. However, published data are mainly based on the OMB racial/ethnic categories. Some have also pointed to difficulties in data analysis that the option for selecting more than one race poses [38].

The use of race and ethnic categories in a single racial/ethnic classification seems to be of common usage, as is the case in the UK census and the Canadian census, denoting a common lack of clarity in the distinction between these two concepts. As presented, since 1980, the census has used the ethnic category "Hispanic" to classify people of Latin American origin. Brazilians, who are of Portuguese rather than Spanish origin and culture, hardly identify with the term "Hispanic" The new OMB terminology added the term "Latino" to "Hispanic", but again, Brazilians rarely identify themselves as "Latinos". The official Canadian classification, for instance, uses the term Latin Americans, a more inclusive term, and of more common usage in the American Continent outside the USA.

"Hispanics" or "Latinos" represent a very heterogeneous group in relation to nationality, race, and even ethnicity. Candace Nelson and Martha Tienda [29] argue that ethnicity is structured by the relationship between a particular national origin group and the organization of society, and that this relationship is shaped over time by immigration history, reception factors when arriving in the US, and race. They have shown large variations between Cubans, Mexicans, and Puerto Ricans in the US in relation to assimilation, residential segregation, and social mobility. Therefore, the self-classification of US residents from Latin American countries depends on people's SES background, time, history, country of migration, and even skin color.

Although less a criticism of OMB standards, the use of "American" in standard terminology for race and ethnicity in the US can be viewed as disrespectful to other countries of the Americas. The use of the adjective "American" to denominate people with origins from other continents or countries (e.g., African American, Italian American, Mexican American, or Brazilian American) disregards the fact that an American is anyone from the American continent, as a European is anyone from a European country or an African is anyone from an African country. Strictly speaking, Brazilian American is synonymous of Brazilian, not a US resident/citizen of Brazilian origin.

The Brazilian experience

Brazil is the country in the Americas that has the largest population of African ancestry [39]. Miscegenation is part of Brazil's history, as in other Latin American countries [40]. The official term for the admixed population in the census is "*Pardo*", literally meaning "Brown" or "Gray". The 2000 census showed that with a total population of 169 million, 53.7% of Brazilians self-assessed themselves as "White", 39.1% as "*Pardo*", 6.2% as "Black", 0.5% as "Yellow", and 0.4% as "Indigenous" (http://www.ibge.com.br, accessed on 10/Feb/2002).

It is important to note that racial measurement in the Brazilian census has always referred to skin color. It refers to phenotype (physical appearance) and not to ancestry (origin), as in the US. Racial categories were created based on a combination of physical appearance and social position [41]. While the "One-Drop" rule shaped the racial division in US society based on "pure" racial categories, the Brazilian census categories have always included a term ("*Pardo*") for the admixed population. Miscegenation was an early pattern in Brazilian society, and the first census in 1872 showed that 38.3% of Brazilians were already mixed ("*Pardo*") [42].

Similar to the US, in the 19th century the gathering of population data in the Brazilian census distinguished free people from slaves and various censuses applied different race/color classifications. The first two general censuses in the country (1872 and 1890) included questions on race. In these censuses, the terminology used was race and not color. In the 1872 census, race was presented in four categories ("White", "Black", "*Pardo*", and "*Cabloco*"). "*Cabloco*" referred to indigenous people and their descendents. In 1890, "*Mestiço*" replaced the term "*Pardo*". From then on the variable was referred to as color until the 1990 census, when color/race was used.

This information was absent from the 1900 and 1920 censuses and was reintroduced in

1940. There was no census in either 1910 or 1930. In 1940, the interviewer, in an open question, made a judgment about the respondent's color. The main options were "White", "Black", and "Yellow". All other designations, indicating admixed individuals (*"Caboclo", "Mulatto", "Moreno"*), were classified under *"Pardo"*. However, official tabulations presented "Black" and *"Pardo"* in aggregate form. In the 1950 census, the major change was that color became self-assessed. Besides, in the questionnaire it was explicitly stated that the term *"Moreno"* could not be accepted as a skin color option.

The 1960 census included the term *"Índio"* [*Indian or Indigenous*], which was previously counted as *"Pardo"*. In this case, *"Índio"* was limited to individuals living on reservations. The question format changed to a five-fold one, but, the micro-data on race from this census were never made available, and published data showed "Indians" collapsed into *"Pardos"*. Data on race were again absent from the 1970 census. This happened during the military dictatorship that ruled the country for more than 20 years and was responsible for violent repression of political groups and severe censorship.

When race was reintroduced into the 1980 census data, the self-classification adopted was four-fold, which again excluded the category "Indian". Eventually, the 1991 and 2000 censuses readopted a five-fold classification, which included "Indigenous" as a separate group. Strictly speaking, "Indigenous" is not a "skin color" category. "Indigenous" might represent some physical traits, ancestry/ethnicity, or even group identity. As such, the present Brazilian census color/race classification also mixes skin color with ethnicity. The color/race question is only included in the census questionnaire applied to a sample of the Brazilian population. Telles & Lim [43] point out that in Brazil, for various reasons, interviewers do not always follow the instructions, and census data on color/race are a combination of self-classification and interviewer classification.

The characteristics of miscegenation in Brazilian society are discussed on the basis of two dominant approaches. Gilberto Freyre, one of the most important and controversial Brazilian sociologists and anthropologists, regarded miscegenation as the core of Brazilian identity [44]. During colonial times there were numerous admixed and illegitimate children of slave masters and priests. According to Jessé de Souza (http://www.iuperj.br/professores/texto3jesse.htm), interpreting Freyre's ideas, the specificity of Brazilian miscegenation is represented *"by the uncertain but real possibility of identifying the patriarch with his illegitimate or natural children with slaves or natives (a Moorish influence, according to Freyre). …We know that the Portuguese, despite being intensely Christian (and even more, champions of the Christian cause against that of Islam), imitated the Arabs, or Moors, the Mohammedans in certain techniques and in certain costumes, assimilating countless cultural values from them. The Mohammedan concept of slavery, as a domestic system linked to the organization of the family, including domestic activities, without being decisively dominated by an economic-industrial proposal, was one of the Moorish or Mohammedan values that the Portuguese applied to colonization, predominately, but not exclusively Christian, of Brazil."* He argues that *"because of the emphasis placed on purity of origin in the United States, this was not even a possibility there"*.

Nineteenth-century urbanization was an opportunity for social mobility for *mestizos*, people that where somehow outside the polar relations established by the master/slave positions. Miscegenation increased substantially in the 19th century, with the proportion of admixed individuals increasing from 10 to 41% of the population (http://www.iuperj.br/professores/texto3jesse.htm).

However, at the turn of the 19th century, miscegenation became related to the notion of *"embranquecimento"* or "whitening" [42]. What became known as the "whitening ideal" aimed to dilute "Black" blood in the "White" blood of European migrants to make the population lighter-colored [42]. The late 19th and early 20th-century Brazilian elites and intellectuals approached "purification" of the population through racial miscegenation and migration, contrary to the US approach to racial "purification" by decreasing the proportion of admixed individuals with "dark color" in the population, using legal enforcement. In contrast, racial discrimination in Brazil has been considered illegal since the beginning of the Republic in 1890 [44]. David Cleary [45] (p. 6) points out that *"the 'whitening thesis' was not unique to Brazil: it was a standard response of Latin American intellectuals to eugenicist orthodoxy, with local variances in Venezuela, Mexico, Argentina, and elsewhere. The contradiction here, that increasing diversity and miscegenation could result in racial uniformity over time, was typical of the mental gymnastics intellectuals of the period were obliged to perform in order to reconcile the tenets of scientific racism with Latin American patent multiracial reality"*.

Three to five million Europeans migrated to Brazil in the late 19th century. The "Euro-

peanization" of Brazil, the respective modern values attached to it, and the "whitening" policies defined the new basis of Brazilian social/racial inequalities. In this new context, "White" was superior to "Non-White", "*but 'White' was (and continues to be) more an indicator of existence of a series of moral and cultural attributes than skin color*" (http://www.iuperj.br/professores/texto3jesse.htm).

There is a strong ongoing debate regarding color/race classification and terminology used by the Brazilian census. The Brazilian Census Bureau (IBGE) organized a Committee to advise the design of the 2000 census, but despite the existing opposition, the final decision by IBGE was not to modify the 1991 classification, maintaining a forced question with a five-fold category.

Color/race classification has been criticized from contrasting viewpoints. The use of discrete race/color classification in a country with a large admixed population has been questioned by various scholars, along with pressure from some scholars and activists [44] for the IBGE to limit its classification to the dichotomous "Black" and "White" US racial approach, by collapsing the category "*Pardo*" with "Black".

Influenced by the experience of other countries, in particular the US, Brazilian scholars and activists of the Black Movement argue in favor of using the dichotomous "Black"/"White" classification. An increasing number of Brazilian activists and scholars, influenced by racial movements in the United States, do not accept that Brazilians be seen as a continuum of colors [44]. The main argument in this case is political: that the emancipation of people with African ancestry requires greater racial polarization than exists in Brazil [46]. Moreover, it is argued that affirmative action policies require identifying who will benefit from them and that the US notion of race as ancestry is more appropriate. However, even in the US, other criteria have been used for affirmative action purposes. Most of the initial affirmative action polices in the US included women as well as racial minorities. And women, most of whom were white, experienced greater economic improvement than minorities in the last three decades [47]. Moreover, US affirmative action policies covered multiple racial/ethnic groups, including "Blacks", "Hispanics", and "American Indians". Thus, affirmative action polices *per se* do not require any particular racial categorization. What is needed is a political decision about which groups should be covered by the program.

It is now becoming common practice among many scholars, activists, policy-makers, and journalists in Brazil to assume the term "*Pardo*" as "Black" to refer to the Brazilian population of African ancestry. "*Pardos*" are aggregated with "Blacks" under the latter designation. Brazilian Ministry of Health documents began to adopt the term "Black" to refer to Brazilians of African ancestry [48]. The same is becoming commonplace in newspapers and scientific articles about racial inequalities [49]. Lovell & Wood [32] used this procedure, but justified it by the fact that the category "Non-White" is more stable over time than "*Pardo*" and "Black", and that the category "White" shows greater reliability. However, it is uncommon for authors to base this decision on technical grounds; in most cases it is a political decision, and not always made explicit.

An opposite view is expressed by other Brazilian and US scholars who recognize miscegenation as part of Brazilian history and consider Brazilian racial identity intrinsically ambiguous [50]. In this case, racial classification does not fit into a discrete "Black"/"White" dichotomy.

Ambiguity is reflected in the fact that Brazilians, when inquired about their color/race in an open-ended question, may answer with 135 to 500 different race-color terms [51,52,53]. *Difícil dizer* [hard to say]; *tostada* [toasted]; *leite* [milk] are some examples, besides the many different terms people use to specify very fine color nuances such as *morena café* [coffee-colored *morena* or tan], *morena clara* [light *morena*], or *morena canela* [cinnamon *morena*]. Ambiguity also refers to the fact that skin color is a very fluid measure that varies greatly with the context [52]. Lovell [42] attributes the influence of social class to the fluidity of color/race identification in Brazil. Some authors say that in Brazil, "money whitens". Wealthier people with darker phenotypes tend to classify themselves and be classified by others in lighter categories [43]. Other contextual circumstances, such as dressing and social status, can also influence color/racial labeling [51]. Given this ambiguity and fluidity of color/race for Brazilians, some have argued that to restrict respondents' choice to a few color/racial categories in the Brazilian census is a violation of the principle of self-identity [51].

Another aspect of skin-color ambiguity is illustrated by the fact that terms such as "*Moreno*" or "*Pardo*" refer to a wide spectrum of phenotypes. According to the standard Brazilian Portuguese dictionary *Aurélio*, "*Moreno*" derives from Spanish and refers to Moorish skin color. Variation in phenotype is even wider for "*Moreno*" than for "*Pardo*" [51].

The usage of the term "*Pardo*" is also highly questioned. Despite the fact that it is adopted

in the census, the word "*Pardo*" is rarely used by Brazilians to refer to skin color. Before the reintroduction of color/race category in the 1980 census, IBGE included in the 1976 PNAD both an open-ended question and a forced-choice question with the color/race categories used in the previous census, to test new ways of inquiring about color/race. In the open-ended question, only a few respondents (6%) recognized themselves as "*Pardo*" [53]. Various studies have found that the most preferred skin color term in the country is "*Moreno*" [51,53]. In agreement with other surveys [51,52], a study published by the *Folha de São Paulo* newspaper [53] showed "*Moreno*" to be the term Brazilians prefer to identify themselves. However, for some [54] "*Moreno*" means both a color and the absence of color.

Nelson do Valle Silva [55] has argued in favor of the validity of the official classification. The core of Silva's argument in favor of the official category is that "*Pardo*" refers to an objective, more precise, "demographic" characteristic, while "*Moreno*" is related to skin-color identity. By doing so, he assumes the variable color/race as being an objective physical trait, not a social construct. According to Silva, for census purposes the demographic trait "color of the skin" is a more appropriate measure than color identity. However, "skin color" matters if it affects peoples' identity in a different manner. It matters when it serves as a determinant of people's choices and opportunities due to discrimination in relation to their phenotype. As a social construct, regardless of the terminology applied or the way the color/race question is constructed or assessed (self-assessed or assessed by the interviewer), skin color will reflect respondent/societal values. It is not a characteristic like age, which despite having social meaning, is an objective and meaningful demographic trait in itself. Moreover, respect for individual dignity is one principle that has to be used in the assessment of color/race. This means that when collecting racial data the most preferred terms that the group uses should be adopted [13].

It should also be noted that the definition of the Brazilian census color/race categories does not support the assumption that "*Pardo*" refers strictly to African descendents. Parra et al. [19] using a classification of color ("White", "Intermediate", and "Black"), based on multivariate evaluation using skin color on the medial surface of the arm, hair color, and texture and the shape of the lips and nose, indicated that for Brazil these phenotypical traits are weak individual predictors of African ancestry estimated on the basis of molecular markers.

Applying an index for African ancestry, they did not find statistically significant differences in their sample in the index for "Blacks" and "Whites". These authors emphasize the risks of equating color with geographic ancestry and of interchangeably using the terms "White" and "European" or "Black" and "African" in the Brazilian context.

Besides inaccuracies regarding ancestry, Brazilian census classification never included a specific category for admixed people with Indigenous ancestry and, as previously shown, until 1980 census data classified all admixed people as "*Pardo*". In the present color/race classification, people of indigenous ancestry can choose to identify themselves as "*Pardo*", "White", or "Indigenous". People self-classified as "Indigenous" or "*Pardo*" in the census forced-choice question tend to classify themselves in similar color/race categories when given the choice to do so. In a recent study [52], comparing census categories with an open-ended question, 61.73% who identified themselves as "Indigenous" in the forced question answered "*Moreno*" in the open-ended question. Similarly, 54% of "*Pardos*" also identified themselves as "*Morenos*". On the other hand, only 12.8% of the people who identified themselves as "Indigenous" in the closed question chose the same option in the open-ended one.

Data from the 1998 National Household Survey (PNAD) signal significant demographic, socioeconomic, and geographic variations across the categories "*Pardo*" and "Black", arguing against aggregation of these two groups. In the North and Northeast, the poorest regions in the country, the majority of the population are "*Pardo*" (68.49% and 64.3% respectively), while in the Southeast and South (the country's wealthiest regions) the majority are "White" (64.03% and 82.91% respectively). The proportion of "Blacks" is higher in the Southeast (7.31%), followed by the Northeast (5.74%). The North is the region with the lowest concentration of "Blacks" (2.2%). People that declared themselves "Yellow" are on average the oldest amongst Brazilians (37.44 years), followed by "Whites" and "Blacks" (29.75 and 30.79 years, respectively). "Indians" and "*Pardos*" are the youngest in all regions (25.51 and 26.16 years, respectively). Among individuals 10 years old or greater, "Blacks" have the worse educational achievement, especially in the Northeast (4.74 years on average; median = 3 years). However, the concentration of "*Pardos*" among the poor is greater than "Blacks": the poorest quintile of per capita family income consists of 23.7% "Blacks" and 30.7% "*Pardos*"

According to Elza Berquó [56], "*Pardos*" and "Blacks" varied in relation to infant mortality, marriage, and fertility. In relation to the above indicators "*Pardos*" were becoming more similar to "Whites". Compared to "Whites" and "*Pardos*", "Blacks" presented higher mortality, later marriage, and higher single marital status, especially among women. From 1960 to 1980, "Black" women presented the highest fertility rate.

In short, collapsing "*Pardos*" into the "Black" category is a questionable approach for measuring people with African ancestry. As shown previously, color is not a reliable marker of genetic variation in the Brazilian population and "*Pardo*" is not a valid category for admixed Brazilians with African ancestry. The simple aggregation of "Blacks" and "*Pardos*" is likely to misestimate the size of the country's African descendent population. Moreover, relevant variations in the demographic and socio-economic characteristics of color/race groups show that these may not be comparable groups and that aggregation may not be the appropriate analytical procedure. Finally, this approach does not respect the color/race category individuals have chosen to self-classify.

Ethnicity is unlikely to be a reliable alternative for skin color/race classification in Brazil. A recent survey conducted by IBGE and representative of the population from six greater metropolitan areas in Brazil allowed the analysis of how Brazilians respond when inquired about their ethnicity. Despite being a society of migrants, Brazil does not show clear cultural divides. When exposed to an ethnicity question with 12 categories of origin, 86% of the respondents identified themselves as Brazilians. Brazilians do not understand the term "origin" in a homogenous fashion. It can mean city of origin, race, or nationality. Responses also varied in relation to age, country/ethnic origin, and time of migration [52].

Given their differences in history and the social meaning of race, in the early 21st century Brazil and the US continue to represent contrasting experiences in relation to racial/ethnic classification. However, the current contrast has a different basis from that of the 19th century. In Brazil, it is represented by the movement to change the racial/skin color classification to an approach based on ancestry, which does not allow for mixed categories. Meanwhile, the increasing intermarriage and migration in the US may soon lead this country to adjust its racial/ethnic classification to a more admixed society for Whites and non-Black groups. Blacks remain distinctive in the US with regard to the persistence of socioeconomic disadvantage and their social and geographic isolation from Whites [14].

These are controversial and politically difficult challenges. Despite the pressure for a more clear-cut classification, Brazilian racial history does not perfectly fit the US racial divide, and many Brazilians apparently resist being classified in discrete and finite "Black" and "White" racial categories. On the contrary, the US population does not easily absorb the mixed-races concept because of the "One-Drop of non-White Blood" rule which constrained the creation of a "multiracial" consciousness. A recent survey showed that only 2% of the US population selected two or more races when given the option to do so [37], which is most probably an underestimation of the multiracial nature of the US population [57]; meanwhile, the term "multiracial" was not clearly understood by all respondents, many of whom preferred not to identify themselves with a multiracial category [37]. However, there are signs that racial consciousness might be changing in the United States. A recent study of racial self-identification by multiracial adolescents showed that this group has a new understanding of race based on the acceptance of diversity and multiracial admixture that does not reflect the influence of the "One-Drop" rule, still prevalent among their parents [58].

In summary, the comparison of race measurement in Brazil and the US shows that the meaning of this concept differs between the two countries. Differences reflect the countries' history, culture, bureaucracy, and political forces that shape color/racial identity. Racial identity is mainly related to how people perceive themselves (and perceive how others perceive them) in relation to their color/race. The concept of race (ancestry) is not the same as skin color. Race in the US originated in the old notion of race as biology, while skin color, despite being associated with racial ancestry, absorbs the social context more and is more fluid. Racial classification is not directly comparable or transposable from one country to another. To measure and interpret race relations in Brazil through the prism of the US experience, or vice versa, is to blind one's analysis of the existing differences, to separate the measurement from its meaning, and to ignore the singularities of each society's historical, cultural, and political experience, which shape people's perceptions, societal values, and the measurement of race itself.

Quality of data

The definition of race is not consistent across societies, as seen in the US and Brazilian examples, nor is it consistent within societies or over time. The difficulties linked to defining race in a precise and unchanging manner directly affect the validity and reliability of this concept. Given the lack of scientific support for the claim that the human species is subdivided into different lineages (races), it is neither possible nor desirable to establish a valid measurement of race in biological terms. At the same time, race is a very meaningful social category that can determine differential access to a broad range of societal resources. Measuring race is not simple or easy because the concepts of race, skin color, ethnicity, origin, ancestry, nationality, and identity overlap. In Canadian statistics there is a preference to use the term ethnicity, but its definitions and measurement represent a clear example of the lack of boundaries between these concepts.

Definitions of racial categories vary across countries and time. Some people that refer to themselves as "Whites" in the US are not similar (in regard to ancestry or skin color) to "Whites" in Brazil. For example, because of the "One-Drop" rule, individuals with "White" skin color but African ancestry are likely to identify themselves as "Black" in the US, but might regard themselves as "White" in Brazil. Even for countries that base racial classification on ancestry or ethnicity (origin) the definition of racial categories is not comparable. "Whites" in the US are not comparable to "Whites" in the UK. The term "White" in the UK never considered Asian Indians, Middle Eastern, and North Africans, but until recently people from India were considered "Whites" by the official US racial classification, and Middle Easterners still are [59]. There are large variations in racial taxonomy across countries. Some taxonomies emphasize ancestry (e.g., US), others ethnicity (UK and Canada), and still others skin color (Brazil). Strictly speaking, these taxonomies, although they overlap in some cases, are not identical.

The understanding of the race concept also varies across respondents and for the same person (on different occasions); the impact of this variation is not the same across racial categories. Large changes in the number of "American Indians" between the 1960 and the 1990 US censuses may be due to improvement in the quality of data, but most likely represent a shift in self-identification. Evidence of a shift in favor of "American Indian" identification ex-

ists and seems to be more common at younger ages, but does not vary by gender. Reasons for changes in self-assessment may be optional or 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*" [25] (p. 174). The same can be true for any other racial category, e.g., the case of US residents from Latin America, as discussed earlier. The proportion of the Brazilian population in each color/racial category also depends on how such categories are presented and interpreted. The meaning of color/racial terminology apparently varies across geographical regions in Brazil [56]. An increasing willingness by some groups of Brazilians to self-identify as "Black" may be the explanation for the relative increase in the proportion of "Blacks" in the 2000 Brazilian census. "Blacks" in the 1991 census [42] represented 5.0% of the Brazilian population and increased to 6.2% in 2000.

As noted previously, the skin-color concept is more influenced by socioeconomic position. Unlike race as ancestry, which bases its definition on the idea of human subspecies, skin color (influenced by the 19th century idea of race) is less rigid and more influenced by context. Labeling and self-assessment of race/color in Brazil has been shown to vary with socioeconomic position. Studies on socio-economic distribution by color/race must be mindful of the fact that color/racial discrepancies may be partially due to the fact that SES shapes color/racial identity and can lead to errors in association when data on race and SES are collected at the same point in time.

It has been suggested that skin color designation in the US is also influenced by socioeconomic position [35] and that skin color within the existing racial categories affects people's life chances (for example, lighter-skinned "Blacks" and "Latinos" experiencing better life chances than those with darker skin). In a national study of "Blacks" in the US, lighter skin color was a stronger determinant of adult income and occupational status than was parental SES. This association was stronger for women than for men [59,60]. As in Brazil, during slavery in the US, "Mulatto" slaves enjoyed more prestige than darker ones [60]. However, unlike a study in which skin color was associated with racial discrimination [60], Krieger et al. [61] failed to find any association between skin color and 5 of 7 specified situations of racial discrimination (getting a job, at work, purchasing a home, get-

ting medical care, in a public setting) in a US sample of young "Black" men and women.

Terminology is another source of disagreement between respondents, who vary in their preferred racial terminology. For instance, in a national study in the US, 44% of "Blacks" preferred "Black", while 28% preferred "African American" [25]. Individuals may also experience difficulties in understanding the terminology of racial classification [62]. This influence is important because variation in terminology across data sources or in time impact the size of the racial groups. As indicated earlier, Brazilians prefer the term "*Moreno*" to "*Pardo*". Harris et al. [52] have shown that the proportion of "Whites" and "Blacks" diminishes significantly when "*Moreno*" is used instead of "*Pardo*" in Brazil.

Self-assessment, interviewer's observation, health provider reporting, and other factors represent different ways of inquiring about race, and each method may influence the response in a different way. It has thus been recommended [13] that researchers specify the way in which data on race by obtained. A study by the US Bureau of the Census showed high agreement between observer and self-assessment responses for "Whites" and "Blacks", but low agreement for the other racial categories [63]. Telles & Lim [43], using 1995 data from a national survey to compare self-assessment of skin color with interviewer classification in Brazil, found that the interviewer classification darkens the skin color of low-SES individuals and whitens it for high-SES individuals.

It is generally recommended that self-assessment is the appropriate way to inquire about race. However, the method should be selected in a more flexible way in relation to the research question. In public health research, observer assessment may be the best option when investigating racial disparities in health care (treatment), because it captures how others perceive the health system client. In contrast, self-assessment may be a more appropriate method for studies on health disparities and disparities in access to health services.

Respondents also show large disagreement between "color/racial" terminology and "ethnic" categories. In the 1980 US census, 26.5 million individuals in the US self-identified themselves as "Black or Negro", but only 21 million reported having African ancestry. Disagreement varies across population groups and was greater for "American Indians" [64] In Brazil, discrepancies are even sharper than in the US, mostly because the majority of Brazilians tend not to identify themselves with distinct origins. In a country with a large admixed

population only, 2.1% identified their origin as "African" and 6.7% as "Native Indians" [52].

Ethnicity may also not be perceived as important when people feel themselves as part of another group. However, when individuals have multiple ethnicities, the choice of one may be situational. Not surprisingly, ethnicity has been found to be quite fluid, with only 58% of people reporting the same ethnicity between two surveys in the US [31].

Racial categories are also very heterogeneous [13]. Particularly due to recent migration, "Black" in the US is increasingly diverse in terms of "ethnic" origin and needs [65] There is also considerable regional variation within the Black population. Scottish-born and Irish-born people living in England and Wales are subgroups of the "White" population with greater needs than some racial/ethnic minorities [59]. Data from the 1998 PNAD show that "Whites" in Brazil are the skin color/race group with the greatest income inequality. Poor "Whites" are far from the rich "Whites" in their needs and lack of opportunities.

The optimal measurement of race depends on the purpose for which race is being used. Within and between countries differences in racial measurement may arise due to diversity in the race concept, data collection methods, question format, terminology, or classification system [60]. Validity and reliability are major problems when measuring race, and measurement error is likely to increase in more interracially or inter-ethnically admixed populations.

Measuring race in admixed populations

The question of whether populations of mixed origins can be categorized into any simple, finite, discrete categories is becoming central to racial/ethnic taxonomy. Some societies have large proportions of admixed people and many others are increasingly becoming admixed. Immigration in the US, especially from Latin American countries, increased in the last few decades, making its population much more heterogeneous. The projection of the US Census Bureau is that by 2050 one half of the US population will be "Non-White" [63] and 21% of the population will be of multiple ancestry [57].

Despite the possibility of answering questions with multiple races, the new OMB classification in the US is not a good solution for classifying admixed people. For miscegenation that goes back many generations, individuals simply do not know about their ancestry. Whenever people's parents, grandparents, and great-grandparents descend from intermarriages of

admixed people, "pure" ancestry becomes very difficult to trace. In Latin American countries such as Brazil where miscegenation occurred at very early stages, it is difficult for a large number of people to answer questions about their origins.

It can also be argued that people do not know their ancestry because origin played a distinct role in societies with early miscegenation. As a result, many people may not find a place in any of the selected discrete "races" categories. In the 2000 US census, 43% of people that identified themselves as "Hispanic or Latino" chose, in the race question, to answer "some other race" (http://www.census.gov/mso/www/rsf/racedata/sld008.htm, accessed on 10/Oct/2002). And they usually inserted their country of origin or an alternative term for their Hispanic ethnicity for their race.

When assessing their own race, recent immigrants from countries where race is not as central in social structure as in the US may apply criteria adopted in their original country. On the other hand, descendents of migrants are more likely to respond to the race question using different criteria from the ones used by their parents. The fact that this classification is based on pure-race categories of ancestry and the absence of a multiracial category increases the chance of misclassification or non-specification for admixed people. On the other hand, multiracial categories tend to be very heterogeneous, and the greater the admixture in a population, the lower the discriminatory power of racial classifications.

Therefore, fluidity and ambiguity of racial measurement increases as the population becomes more multicultural and admixed. The more admixed a society, the greater the misspecification and heterogeneity of racial categories based on ancestry. Bias will also affect classifications that allow people to be classified in more than one pure-race category, as in the new US classification. Multiracial categories also tend to be very heterogeneous. At the same time, US data on children born to Black/White unions indicate that infants with a Black mother and White father consistently have higher health risks than those with a White mother and Black father [65], suggesting that in at least some situations there may be health risks linked to the specific pattern of multiracial status.

The use of skin color may be a more adequate proxy for racial/ethnic discrimination in admixed societies than racial measurement based on ancestry. Ethnicity or nationality may also be more meaningful in societies with recent migrants.

Racial measurement in health-related documents

Validity and reliability problems with racial measurement are of major concern in public health. The various quality problems associated with race data lead to bias in health indicators. Bias can happen whenever individuals in the numerator do not come from the same population as in the denominator. Inconsistencies also vary between indicators and are expected to be greater for admixed populations. Problems of comparability increase when rates are used for the purpose of between-country comparisons, given differences in racial, conceptual, classificatory, terminological, and data-collection methodology.

Racial classification, question content and format, and data collection methods, which impact racial estimates, are not and cannot always be the same across data sources. In general, health indicators depend both on census/survey and vital statistics data. In routine health indicators, censuses and surveys frequently provide data for the denominator, while data for the numerator come from a different source. In this case, data are collected at different points in time, potentially affecting the classification. As discussed before, people change their racial (ethnic or color) self-assessment over time and in relation to context, introducing variation in health indicators that may not be related to variation in frequency of the event under study, but to the size of the population in the race category due to changes in the way people identify themselves. Although there is little evidence that this currently distorts US racial data on health, it is likely to be of increasing importance in the future due to increasing migration and intermarriage. This variation in assessment will also affect time-trend analysis. Data collection methods can also vary. Self-assessment is generally used in census and household surveys, but it is not always possible to be ascertained on death certificates, for example. Data on race based on information from medical records (which are more likely to be obtained by hospital staff or physicians) are likely to vary from data obtained by census interviewers.

As noted previously, in Brazil data on color/race only became mandatory on death certificates and in the SINASC in 1986 and were also included more recently on the forms for compulsory notification of diseases (SINAN) and in protocols for research involving human beings. Reporting of this variable increased year by year, with about 12% of missing data in both

mortality and SINASC data in 2002. Health services data systems do not include information on race. The color/racial classification is the same used in the census, but data collection methods are not standardized. On the death certificate, the physician generally informs the color/race of the deceased, but in case of in-hospital deaths some other hospital staff members can provide this information. The same procedure is adopted for deaths of children. In the SINASC, the mother reports the color/race of the newborn. Estimates of infant mortality rates by color/race based on SINASC data may be particularly affected by differences in data collection methods between the death certificate and SINASC. This can also have an impact on life expectancy estimates based on color/race.

The US experience in measuring race on birth and death certificates demonstrates some of the problems with racial measurement on health documents. Until 1988, the National Center for Health Statistics used a complex algorithm for birth certificates which assumed "White" as the only race that required both parents to be "White". Based on the "One-Drop" rule, these criteria implied that when one of the parents was "White" but the other was not, the child would receive the race of the "non-White" parent. This changed in 1989, and the United States now reports infant statistics by the mother's race. However, the definition on the birth certificate is not always the same as on the death certificate. On the death certificate, funeral directors (or other certifiers) are supposed to ask the next-of-kin to inform the deceased's "self-assessed" race. Most funeral home directors do not follow this procedure. Accordingly, race on the death certificate has significant inconsistencies with racial data in the census for groups other than Black or White, thereby affecting death rates. Inconsistency in racial measurement on birth and death certificates for children born between 1983 and 1985 in the US and who died within one year was 43.2% for all races [66]. This problem also exists on death certificates of adults, with many individuals who would have self-identified as "American Indian", "Asian", or "Hispanic" being misclassified as non-Hispanic White on the death certificate. This numerator problem understates the nationally reported mortality rates for these groups.

In the US, problems have also been documented with the denominator. Estimates of an under-count of middle-aged Black men in the census lead to the overestimation of mortality rates (and other health data that use census da-ta for the denominator) for this population group. It is also assumed that underestimation on the census data of this group is higher in some urban areas [65]. A recent study that reviewed race and ethnicity data in US Public Health Data Systems in New England concluded that data from these systems are inconsistent, unclear, and not comparable. Moreover, some classifications presented error in relation to geography and social/cultural realities of race/ethnicity categories [67]. The New England area has a large concentration of Brazilian residents and descendents, but the classifications used are not appropriate to classify Brazilians in a single race, given the differences in racial identity and the large number of admixed people.

In short, lack of accuracy on race-specific health indicators is directly related to reliability problems in racial measurement, and there are no easy solutions for dealing with many of the possible errors. Errors lead to over- or underestimates, and the direction and magnitude of these miscalculations vary across race categories, health indicators, and time. Differences in the definitions of race and reliability problems also point to low internal and external validity of race-related studies in public health.

Given these limitations and to avoid the misuse of race, ethnicity, and nationality as biological variables, various scientific journals, including *Paediatric and Perinatal Epidemiology* [68] and *Nature Genetics* [69] recently issued rules for articles to be published using any of these variables. In a recent Editorial in *The New England Journal of Medicine* [70] (p. 1393), the author makes similar claims stating "*as for medical research, any investigator that entails so-called racial distinctions, whether a clinical trial or a laboratory study, should begin with a plausible, clearly defined, and testable hypothesis… but, tax-supported trolling of data bases to find racial distinctions in human biology must end*".

Why keep on measuring race?

Race is a 19th century biological concept that lost its once-claimed scientific support and that is no longer seen as the appropriate way to approach genetic differences across population subgroups. As a social construct, its definition varies across societies, groups, and individuals, and it is challenging to measure race in a precise and reliable manner. Increasing migration and multiracial marriages pose newer and greater difficulties for racial taxonomy.

Some scholars and users in the US, Canada, and other countries claim that race should be

abandoned because it is an ambiguous and vague concept with enormous problems of reliability and validity. During the discussion for review of OMB categories, proposals were made for the complete elimination of racial classification in the US. On the other hand, many authors, despite recognizing these limitations, argue in favor of the use of race. They argue that because race has historically and currently reflected differences in power and desirable access in society, as long as these inequalities exist, race should be measured to monitor them and build societal support to eliminate them [2,14,65].

Racial disparities are the expression of social stratification based on racial discrimination and racial stigma. Glenn Loury [71] (p. 167) points to differences between racism and stigma: racism is related to discrimination in treatment, while stigma "*is about who, at the deepest level, they* [Afro-Americans] *are understood to be*". According to Loury, racial discrimination and racial stigma are distinct phenomena, both rooted in the social context. As a social construct, racial measurement captures social status related to the social history of each society.

Racial discrimination is assumed as one important pathway associated with poorer health in particular racial groups [12]. The other aspect of racial discrimination relates to inequalities in treatment (quality of care), a reflection of health professionals' discrimination. Existing evidence indicates that discrimination is not uncommon in health services [4], and it occurs across societies and is not only focused on race. Various patients' characteristics besides race, such as gender, income, and physical appearance have been shown to be associated with differences in physician decision-making and behavior [72]. Discrimination and stigma are aspects of human behavior that deleteriously affect the health and quality of care received by socially disadvantaged groups. Understanding discrimination and the relationship between discrimination and health, access to services, and treatment in a broader way is important. These represent a complex phenomenon that requires clear understanding of the specificities of stigma and discrimination for allowing interventions to reduce inequalities that are not narrowly focused.

Crude measures of racial disparities in health and health care cannot be assumed as the existence of a causal relationship between race and health. Race as a *risk marker* is not synonymous with race as a risk factor [73]. Increased frequency of a health indicator in a given racial group is informative for policy-making, but it should not be directly assumed as a causal relationship. Race must be studied and understood within the context of other relevant social inequalities. Social position and discrimination have been shown to affect health, access, and quality of care in a complex and cumulative way [74], implying that the analysis of racial disparities in health and health care should take these relationships into consideration to avoid spurious association between health and race. When comparing racial differences, one has to adjust for all relevant SES measures assumed as potential confounders, but residual confounding may lead to a spurious independent effect of race [75].

Finally, since race is a proxy measure for other factors besides biology, some authors, such as LaVeist [12], have pointed to the need for finding more creative ways of measuring such factors.

Recommendations

• The use of race in public health data should be justified and racial measurement clearly conceptualized. It is also important to shed light on what race is a proxy for and to provide indications on how findings should be interpreted.
• Health documents should include at least one social variable to allow for the monitoring of social inequalities. However, whenever data on race are included they should be accompanied by one or more social stratification variables to avoid misspecification of complex health risks or harmful stereotyping. The use of race in public health needs to avoid simplistic conclusions and interpretations, which can lead to a spurious salience of race in the explanation of health and health-care inequalities.
• The decision to incorporate race in routinely collected health statistics should consider validity and reliability problems. To increase the quality of data, data collection methods (self-report, interviewer assessment, other) for measuring race should be clearly stated in each document, and the data collection process should be standardized. The data collection method may vary in relation to the objective of the study.
• Race-specific indicators based on data from a single source, like census or survey data, are likely to display greater reliability than racial indicators based on different data sources. Data collection forms should be standardized to ensure consistent classification across data systems.
• The publication of race-specific indicators should be accompanied by information on the

concept of race being used, data limitations, and potential bias to avoid misinterpretation.

• Publication of variations in frequencies in race-related health indicators should specify that they cannot be assumed as evidence for racial causality. Within-group variations should also be published whenever possible. Ideally, publications should include possible explanations for observed variations.

• Aggregation of racial categories, for example "*Pardo*" under "Black", unless justified on the grounds of data homogeneity in relation to other relevant demographic and social variables, should not be an acceptable public health practice.

• The use of skin color measurement in public health should be avoided as a means of identifying people's geographic origin or as a genetic marker. Such data may be useful as a marker of the risk of discrimination or other social exposures.

• Research is needed to test for more acceptable and adequate measurement of skin color in Brazil. This should be associated with the research directed at providing a clearer understanding of the meanings of skin color in the country.

• Since the concept and classification of race are not uniform within datasets, between countries, or over time, any comparison of study results concerning the effects of race on health and health care must take into consideration variations in racial measurement use and data quality issues.

Resumo

Raça tem sido amplamente utilizada em estudos sobre desigualdades em saúde e assistência, principalmente nos Estados Unidos. Diversos problemas relativos à validade e confiabilidade da mensuração de raça são relevantes para o campo da saúde pública. Este artigo faz uma revisão da literatura sobre o conceito e mensuração de raça e compara como os achados se aplicam aos Estados Unidos e ao Brasil. Os autores discutem em detalhes diversas questões da qualidade dos dados relacionadas à mensuração de raça e os problemas levantados pela mensuração de raça em sociedades multirraciais como o Brasil. Além disso, analisam como essas questões se aplicam à saúde pública, e formulam recomendações sobre a mensuração de raça em registros médicos e pesquisa em saúde pública.

Raça; Sociologia Médica; Eqüidade; Literatura de Revisão; Cor da Pele

References

1. Goodman AH. Why genes don't count (for racial differences in health). Am J Public Health 2000; 90:1699-702.

2. Krieger N, Rowley DL, Herman AA, Avery B, Phillips MT. Racism, sexism, and social class: implications for studies of health, disease, and well-being. Am J Prev Med 1993; 9:82-122.

3. Williams DR, Collins C. US socioeconomic and racial differences in health: patterns and explanations. Annu Rev Sociol 1995; 21:349-86.

4. Smedley BD, Stith AY, Nelson AR, editors. Unequal treatment: confronting racial and ethnic disparities in health care. Washington, DC: National Academies Press; 2003.

5. Batista LE. Pode o estudo da mortalidade denunciar as desigualdades raciais? In: Barbosa LMA, Silva PBG, Silvério VR, organizadores. De preto a afro-descendente: trajetos de pesquisa sobre relações étnico-raciais no Brasil. São Carlos: EdUFSCar; 2003. p. 243-60.

6. Cunha EMGP. Mortalidade infantil e raça: as diferenças da desigualdade. Jornal da Rede Feminista de Saúde 2001; 23. http://www.redesaude.org.br.

7. Martins AL, Tanaka ACD. Mulheres negras e mortalidade materna no estado do Paraná, Brasil, de 1993 a 1998. Revista Brasileira de Crescimento e Desenvolvimento Humano 2000; 10:27-38.

8. Dachs NW. Determinantes das desigualdades na auto-avaliação do estado de saúde no Brasil: análise dos dados da PNAD/1998. Ciênc Saúde Coletiva 2002; 7:641-57.

9. Barros FC, Victora CG, Horta BL. Ethnicity and infant health in Southern Brazil. A birth cohort study. Int J Epidemiol 2001; 30:1001-8.

10. Castro MSM, Travassos C, Carvalho MS. Fatores associados às internações hospitalares no Brasil. Ciênc Saúde Coletiva 2002; 7:795-811.

11. Kaufman JS. How inconsistencies in racial classification demystify the race construct in public health statistics. Epidemiology 1999; 10:101-3.

12. LaVeist TA. Beyond dummy variables and sample selection: what health services researchers ought to know about race as a variable. Health Serv Res 1994; 29:1-16.

13. Williams DR. Race and health: basic questions, emerging directions. Ann Epidemiol 1997; 7:322-33.

14. Maio MC. O Brasil no concerto das nações: a luta contra o racismo nos primórdios da UNESCO. Hist Ciênc Saúde 1998; 5:375-413.

15. Littlefield A, Liberman L, Reynolds LT. Redefining race: the potential demise of a concept in physical anthropology. Curr Anthropol 1982; 23:641-55.

16. Malabar FL, Robert E. Anthropological glossary. Melbourne: Krieger Publishing; 1995.

17. Templeton AR. Human races: a genetic and evolutionary perspective. Am Anthropol 1998; 100: 632-50.

18. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. Science 2002; 298:2381-5.

19. Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SD. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 2003; 100:177-82.

20. Rose S. Lifelines – Biology beyond determinism. New York: Oxford University Press; 1998.

21. Cooper RS. Gene-environment interactions and the etiology of common complex disease. Ann Intern Med 2003; 139:437-40.

22. Risch N, Burchard E, Ziv E, Tang H. Categorization of humans in biomedical research: genes, race and disease. Genome Biol 2002; 3.

23. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. Population genetic structure of variable drug response. Nat Genet 2001; 29:265-9.

24. Fuchs SC, Guimarães SM, Sortica C, Wainberg F, Dias KO, Ughini M, et al. Reliability of race assessment based on the race of the ascendants: a cross-sectional study. BMC Public Health 2002; 2:1.

25. Williams DR. Race, socioeconomic status, and health: the added effects of racism and discrimination. Ann NY Acad Sci 1999; 896:173-88.

26. Porter D. "Enemies of the race": biologism, environmentalism, and public health in Edwardian England. Vic Stud 1991; 34:159-78.

27. Marger MN. Race and ethnic relations: American and global perspectives. 6th Ed. Belmont: Wadsworth; 2003.

28. Wright EO. Class counts: comparative studies in class analysis. Cambridge: Cambridge University Press; 1997.

29. Nelson C, Tienda M. The structuring of Hispanic ethnicity: historical and contemporary perspectives. In: Alba RD, editor. Ethnicity and race in the U.S.A. London: Routledge and Kegan Paul; 1985. p. 49-74.

30. Williams DR, Lavizzo-Mourey R, Warren RC. The concept of race and health status in America. Public Health Rep 1994; 109:26-41.

31. Lin SS, Kelsey JL. Use of race and ethnicity in epidemiologic research: concepts, methodological issues, and suggestions for research. Epidemiol Rev 2000; 22:187-202.

32. Lovell PA, Wood CH. Skin color, racial identity, and life chances in Brazil. Lat Am Perspect 1998; 25:90-109.

33. Nobles M. History counts: a comparative analysis of racial/color categorization in US and Brazilian censuses. Am J Public Health 2000; 90:1738-45.

34. Byrd WM, Clayton LA. Racial and ethnic disparities in health care: a background and history. In: Smedley BD, Stith AY, Nelson AR, editors. Unequal treatment: confronting racial and ethnic disparities in health care. Washington, DC: National Academy Press; 2002. p. 248-72.

35. Hill ME. Color differences in the socioeconomic status of African American men: results of a longitudinal study. Soc Forces 2000; 78:1437-60.

36. Byrd WM, Clayton LA. Changing concepts of race and ethnicity. In: Smedley BD, Stith AY, Nelson AR, editors. Unequal treatment: confronting racial and ethnic disparities in health care, Washington DC: National Academy Press; 2002. p. 273-8.

37. Wallman KK, Evinger S, Schechter S. Measuring our nation's diversity: developing a common lan-

guage for data on race/ethnicity. Am J Public Health 2000; 90:1704-8.

38. Harrison RJ. Inadequacies of multiple response race data in the Federal Statistical System. In: Multiraciality: how will the new census data be used? New York: Levy Institute.

39. Torres C. Descendientes de africanos en la Region de las Américas y equidad en materia de salud. Rev Panam Salud Publica 2002; 11:471-9.

40. Sansone L. Racismo sem etnicidade. Políticas públicas e discriminação racial em perspectiva comparada. Dados 1998; 41:751-83.

41. Agier M. Racism, culture and black identity in Brazil. Bull Lat Am Res 1995; 14:245-64.

42. Lovell PA. Development and the persistence of racial inequality in Brazil: 1950-1991. J Dev Areas 1999; 33:395-418.

43. Telles EE, Lim N. Does it matter who answers the race question? Racial classification and income inequality in Brazil. Demography 1998; 35:465-74.

44. Fry P. Politics, nationality, and the meaning of "race" in Brazil. Daedalus 2000; 129:83-116.

45. Cleary D. Race, nationalism and social theory in Brazil: rethinking Gilberto Freyre. Cambridge: David Rockefeller Center for Latin American Studies, Harvard University; 1999.

46. Hanchard M. Orpheus and power: the Movimento Negro of Rio de Janeiro and São Paulo, Brazil, 1945-1988. Princeton: Princeton University Press; 1994.

47. Changing America: indicators of social and economic well-being by race and Hispanic origin. Washington, DC: Government Printing Office; 1998.

48. Ministério da Saúde. Manual de doenças mais importantes, por razões étnicas, na população Brasileira afro-descendente. Brasília: Ministério da Saúde; 2001.

49. Henriques R. Desigualdade racial no Brasil: evolução das condições de vida na década de 90. Brasília: Instituto de Pesquisa Econômica Aplicada; 2001. (Texto para discussão; 807).

50. Motta R. Paradigmas de interpretação das relações raciais no Brasil. Estudos Afro-asiáticos 2000; 38:1-17.

51. Harris M, Consorte JG, Lang J, Byrne B. Who are the whites? Imposed census categories and the racial demography of Brazil. Social Forces 1993; 72:451-62.

52. Schwartzman S. Fora de foco: diversidade e identidades étnicas no Brasil. Cadernos de Pesquisa 1999: 106:83-96.

53. Turra C, Venturini G. Racismo cordial. São Paulo: Ática, 1995.

54. Maggie Y. Aqueles a quem foi negada a cor do dia: as categorias cor e raça na cultura brasileira. In: Maio MC, Santos RV, organizadores. Raça, ciência e sociedade. Rio de Janeiro: Editora Fiocruz; 1996. p. 225-34.

55. Silva NV. Morenidade: modo de usar. Estudos Afro-asiáticos 2002; 30:79-95.

56. Berquó E. The demography of inequality: considerations on Black people in Brazil. Brazilian Journal of Population Studies 1997; 1:165-77.

57. Waters MC. Immigration, intermarriage, and the challenges of measuring racial/ethnic identities. Am J Public Health 2000; 90:1735-7.

58. Harris DR, Sim JJ. Who is multiracial? Assessing the complexity of lived race. American Sociological Review 2002; 67:614-27.

59. Bhopal R, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. Am J Public Health 1998; 88:1303-7.

60. Keith VM, Herring C. Skin tone and stratification in the Black community. AJS 1991; 97:760-78.

61. Krieger N, Stephen S, Coakley E. Racial discrimination and skin color in the CARDIA study: implications for public health research. Am J Public Health 1998; 88:1308-13.

62. Sondik EJ, Lucas JW, Madans JH, Smith SS. Race/ethnicity and the 2000 census: implications for public health. Am J Public Health 2000; 90:1709-13.

63. Mckenney NR, Bennett CE. Issues regarding data on race and ethnicity: the Census Bureau experience. Public Health Rep 1994; 109:16-25.

64. Williams DR. Race/ethnicity and socioeconomic status: measurement and methodological issues. Int J Health Serv 1996; 26:483-505.

65. Williams DR, Jackson JS. Race/ethnicity and the 2000 census: recommendations for African American and other black populations in the United States. Am J Public Health 2000; 90:1728-30.

66. Hahn RA, Mulinare J, Teutsch SM. Inconsistencies in coding of race and ethnicity between birth and death in U.S. infants: a new look at infant mortality, 1983 through 1985. JAMA 1992; 267:259-63.

67. Laws MB, Heckscher RA. Racial and ethnic identification practices in public health data systems in New England. Public Health Rep 2002; 117:50-61.

68. Editorial. Race, ethnicity and nationality. Paediatr Perinat Epidemiol 2000; 14:4-13.

69. Editorial. Census, race and science. Nat Genet 2000; 24:97-8.

70. Schwartz RS. Racial profiling in medical research. N Engl J Med 2001; 344:1392-3.

71. Loury GC. The anatomy of racial inequality. Cambridge/London: Harvard University Press; 2002.

72. Eisenberg JM. Physician utilization: the state of research about physicians' practice patterns. Med Care 1985; 23:461-83.

73. Osborne NG, Feir MD. The use of race in medical research. JAMA 1992; 267:275-9.

74. House JS, Williams DR. Understanding and reducing socioeconomic and racial / ethnic disparities in health. In: Smedly BD, Syme SL, editors. Promoting health: intervention strategies from social and behavioral research. Washington, DC: National Academy Press; 2000. p. 81-124.

75. Kaufman JS, Cooper RS, McGee DL. Socioeconomic status and health in blacks and whites: the problem of residual confounding and the resiliency of race. Epidemiology 1997; 8:621-8.