Stratifying a Population by Race

Michael Root

1. Introduction

Philosophers have written a number of interesting books and articles about race in recent years, but most are about what race is (about the nature or reality of race or racial categories) or about how race bears on issues of distributive or retributive justice. ¹ My interest is different: not in what race is but what race does and, in particular, how race is used in the biomedical sciences to describe or explain differences within the U.S. population in health status or outcomes.

Biomedical scientists in the United States routinely stratify populations by race when describing or explaining variations in health risks and find that Blacks and Whites differ in their risk of a variety of diseases and disease-related deaths.² These findings are important, since they shape public policy and guide efforts to improve public health.³

However, a number of objections have been raised against the use of race as a variable in biomedical research. I discuss three objections and explain how studies of racial differences in the United States can avoid them and how race can be better used in the biomedical sciences. Rather than treat race as a fixed characteristic of members of a population, the sciences, I argue, should treat race as a variable whose value can vary from one health risk to another. Race, on this view, is not a property of members of a population but a relation between members and the risk within the population of morbidity and mortality; while, at the present time, the biomedical sciences assume that a member’s race is fixed, they should assume that his race can vary and that a member’s race can be White in relation to his risk of one medical condition or outcome and Black in relation to another.

2. Objections to the Use of Race

According to one prominent objection to the use of race in the biomedical sciences, there is no way to define “race” and no objective way to assign members of a population to one rather than another racial category. As a result, “race” has no scientific meaning and should not be used as a variable in scientific research or to describe or explain differences in the rates of morbidity and mortality within a population (Witzig 1996).

According to another prominent objection, the groups picked out by our common racial labels are not natural biological kinds or genetically distinct
populations. The common practice of grouping people by race rests on the false belief that there is some genetic property, a race gene or gene cluster, that individuates the races and explains a number of alleged differences in the behavior or anatomy of members of the population. There aren’t such genes. As a result, there are no races (Appiah 1990).

According to a third prominent objection, the use of race as an identifier in social or biomedical research encourages racial stereotypes and racialist thinking. Claims of racial differences in health encourage the view that some races are more fit than others or that racial inequalities in health are due to differences in human nature. Though the sciences might oppose racialist thinking, their use of race to describe or explain differences in health within the human population helps to sustain it (Stolley 1999).

These three prominent objections to the use in the sciences are often used to support the view that racial labels should not be used in studies of human variation and the biomedical sciences should not classify or identify members of the population by race. They are also used to support the view that there is no such thing as race, that race is a myth or an illusion or that race has no scientific meaning. Race, on my view, is not an illusion, and the use of racial labels should continue to be employed in studies of differences within populations in rates of death and disease as long as members of the population practice racial discrimination or base their behavior on the race they take themselves to be. However, changes need to be made in how race is assigned to members of a population within these studies if these objections are to be met and before the affects of race on health can be correctly understood.

### 3. Assigning Race

In order to stratify a population by race, a researcher has to assign each member to a racial category from a list of categories. In the United States, members of the population are usually assigned to one or more of the following five racial groups (those identified by the U.S. Office of Management and Budget (OMB) and required to be included in all official statistics): (1) Black, (2) White, (3) American Indian or Alaskan Native, (4) Asian, (5) Native Hawaiian or Other Pacific Islander (OMB 1997). While the OMB does not say how members of the population are to be assigned to the categories, biomedical scientists make the assignments in one of four ways: by self-reports of race; by other-reports of race; by biological mother’s race (as entered on a birth certificate); or by race on a death certificate.

Self-reported race has become the most common way of assigning race in biomedical research. The U.S. government relies on self-reports when collecting data on differences in health, and most epidemiological studies of racial differences in morbidity or mortality do as well (Friedman, Cohen, Averbach, and Norton 2000). Many epidemiologists call self-reported race “the gold standard” when studying how rates of morbidity or mortality vary between Blacks and
Whites and recommend that, whenever possible, self-reports be used when collecting racial data (Kaufman and Cooper 2001). Assigning an individual the race she assigns herself is often the easiest or most respectful way to assign her to a racial category. By allowing each individual to be the arbiter of her own race, we display the subjective and social nature of our system of racial classification and give individuals control over their own identity. Nevertheless, the easiest or most respectful way for epidemiologists to identify the race of their subjects might not give the categories as much descriptive or explanatory power as a less easy or respectful way.

With many biomedical variables, there is overwhelming evidence that self-reports are not reliable or accurate and that the actual value of a variable is often different from the self-reported one. For example, self-reported body size is not a very accurate measure of actual size (BMI), and self-reports of alcohol intake are not a very accurate measure of actual intake. Self-reports of body size and alcohol intake are often biased and, as a result, do not accurately measure how differences in size or intake affect health. Why then do epidemiologists take self-reported race to be an accurate measure of a member’s race in their studies of differences in morbidity and mortality?

Most epidemiologists assume that in the vast majority of cases each of the different ways of assigning race to a member of a population gives the same result, and, consequently, her self-reported race is a good proxy for her actual race even if her actual race is her other-reported or mother’s race. However, there is evidence that, for an increasingly large percentage of the population, the different ways to assign race do not give the same result. Many studies conducted in the last ten years suggest that in the case of children of mixed-race parents, Hispanics and foreign-born Americans, self-reported and other-reported race are often different (Harris and Sim 2002). In addition, the self-identified race of many members of the U.S. population is fluid rather than fixed; how a member reports her race depends on the choices she is given, who is asking her about her race, and where or how she is asked, and, even when her other-identified race is fixed and others take her to be Black, she might respond that she is White or not to report her race when asked the race question or questioned about her identity (Waters 2000). Other-reports of race and mother’s race are in many cases more stable or robust than self-reports of race and vary less with context. As a result, self-reports are less reliable measures of race of many in the U.S. population than other-reports or mother’s race, and a social or biomedical scientist can have a reason to assign a race to a member different from the one she assigns herself.

Many members of the population have a variety of context-specific self-identities, but, despite their self-identities, if many are consistently identified as Black by others, then a biomedical scientist interested in how access to housing, education, mortgage lending, employment opportunities or rates of death or disease vary with race has a reason to take other-reports rather than self-reports as the best measure of the person’s race, and officials who decide economic or health policy have a reason to take other-reports rather then self-reports as the measure.
of the actual race of members of the U.S. population when they measure or monitor racial differences in economic or health outcomes.

Always equating a member’s race with his self-reported race can result in a biased estimate of the effect of race on morbidity or mortality within the population if the self-reported race of members is not a good proxy for their other-reported race or their race on their birth or death certificate. If mother’s race is a risk factor for X and self-reported race is not, then to stratify the population by self-reported rather than mother’s race in a study of X will underestimate the racial difference in risk of X within the population.

4. Error Measurement

Epidemiologists stratify the U.S. population by race and find that morbidity and mortality rates are higher for one race than another. But are their findings accurate or is there error in their measurement of the race of members of the population? Bias or error in the measurement of a variable is only possible if the variable has an actual or true value from which the apparent value can diverge.

If T is the actual value of the variable, O the measured or apparent value, and e the error in the measurement of the variable, then $e = O - T$, but e has no value unless the variable has an actual value. If beauty were simply in the eye of the beholder, then no measurement of the beauty of members of a population could be biased or in error, and no report could over-count or under-count the number of members who are beautiful, since there would be no difference between a member’s observed and her actual or true beauty. For a measure of the race of a member of the population to be in error, he would have to have an actual race different from his observed one.

According to a study of the quality of mortality statistics published by the National Center for Health Statistics (NCHS) between 1979 and 1989, death certificates misclassified the race of some decedents, and the actual mortality rates for some races were greater than the NCHS reported (Rosenberg, Mauer, Sorlie, and Johnson 1999). The study assumed that each decedent has an actual race and that his actual race is not always the race he appears to be and, in particular, that the race he is assigned on an official document like a death certificate can be different than his actual race.

The death rate for members of a race X is equal to the number of X deaths divided by the number of Xs in the population. Measurement errors can occur either in the numerator or the denominator. Errors in the numerator are measured, as a rule, by comparing the race assigned to an individual on his death certificate with the race he had assigned himself on a population survey before his death or, in the case of an infant, with the race on his birth certificate (usually based on his mother’s race). The decedent’s race in the survey is taken to be his actual race and on his death certificate his apparent one. Were the race on his death certificate taken to be his actual race, then any difference between his race on his death
certificate and the survey or birth certificate would mean that the survey or birth certificate rather than the death certificate was in error.

Moreover, if a decedent’s actual race were his race at death, then the number in the denominator rather than numerator would be in error, and the NCHS report would have understated rather than overstated the death rate for members of his race in the U.S. population. Since the true measure of the rate of death within a racial group depends on how race is assigned to the deceased members, we cannot know the direction or degree of the error in the NCHS report unless we know how race should be assigned to members of a population when preparing morbidity or mortality statistics. Furthermore, unless the deceased members of a population have an actual race, no count of decedents by race or measure of the difference in mortality rates between any two OMB races could be in error.

5. Actual Race

Any population survey or study of health risks identifies an individual’s race relative to a particular interest. In the case of the U.S. Census, a major interest is to monitor civil rights and, in the case of many health-related surveys, a major interest is to monitor differences in access to medical care within a population. Members of a population do not have a race in the sense in which they have a date or place of birth, but they differ in their risk of heart disease or diabetes, and the difference can be the result of the race they take themselves to be, the race they are taken to be by others or their mother’s race.

If race were a natural biological kind, then the true or actual race of a member of the population would depend on his core biological properties or genotype, but race is not a natural biological kind; nevertheless a member’s race is not whatever anyone takes his race be. Race, like marital status, is not an intrinsic property of persons, but, given how race and marital status are assigned in the United States, a distinction can be drawn between being married and pretending to be, or being White and passing as White. However, since, within the United States today, there is not one but four ways members of the population assign a race to one another: a member’s actual race can be equated with his self-reported, other-reported or mother’s race, and his actual race can be in doubt.

Which race is a member’s actual race, on my view, depends on the trait whose variation within the population we are trying to understand. A member’s self-reported race might be his actual race if the trait is obesity, since what he takes his appropriate body size to be is affected by the norms or ideals of those he takes to be his peers and, in particular, the racial group he most identifies with (Lynch et al. 2009). But then again his actual race can be his other-reported race if the trait is his risk of being denied a medical treatment, since if a health provider denies him a treatment based on his race, the race will be the race the provider takes him to be (Schauer 2003). However, his actual race might be his mother’s race if the trait is his risk of an inherited illness like sickle-cell disease (Root 2003).
Other-reported race describes or explains differences in risk between populations better than self-reported race to the extent that the differences are primarily due to a member’s exposure to racial discrimination, since his exposure to discrimination is not based on the category he assigns himself but the one he is assigned by others (Root 2007). As a result, whenever differences between groups in risk are primarily due to racial discrimination, self-reported race should be employed as a standard if and only if there is good evidence that self-reports are a reasonably good proxy for the reports of others and, in particular, those others who are engaged in discrimination.

Race is not only used within the biomedical sciences as a descriptive variable but as an explanatory one as well; race is used to describe how risks of death or disease vary within a population but also to explain why they do. Differences between racial groups in morbidity or mortality are often explained as the result of a difference between them in (a) gene frequencies, (b) cultural practices, (c) access to a material resource, or (d) proximity to an environmental hazard. These explanations are problematic for, given the way membership in these groups is determined, differences in (a), (b), (c), or (d) within each group are often as great as the differences between the groups. Some of the problem has to do with the nature of the categories themselves, since the races, no matter how we define them, are not natural kinds, and the members of a racial group do not share any core properties (in the way that instances of a natural kind like gold or water do). No matter which way we choose to distinguish Blacks from Whites, the variation within each racial group in (a)–(d) will be great. However, the degree of intra-group difference can vary with how race is assigned race to members of the population, and, given an interest in explaining differences within the population in a biomedical trait, the best way to assign race is whichever increases the difference between groups and decreases the difference within groups in the trait, since where there is little or no difference between members of different races or great difference between members of the same race, the categories are of little use when trying to explain a variation in the trait within the population.

As a result, there can be good reason to assign a member of a population to one racial group for a study of variations in one trait, for example, a stress-related illness, and a different one for another, for example, inherited diseases. In the absence of any interest in the influence an individual’s race has on a person’s prospects, for example, risk of morbidity, there is no more reason to take his actual race to be his self-reported rather than his other-reported or mother’s race. Because there is no one way to establish his actual or true race but many, and since each has its advantages and disadvantages, there is no way to decide his actual race that does not depend on the purpose that the decision is intended to serve.

6. Pluralism

At the population level, race is more like poverty or unemployment than date or place of birth. Whether a person is poor or unemployed depends on the standard
or criterion of poverty or unemployment used to assign members of the population to those categories. The value of the variable depends on a standard, but the standard depends on interests. According to the U.S. Bureau of Census, for example, the U.S. poverty rate rose from 11.7 percent in 2001 to 12.1 percent in 2002. The statistics are based on a standard of poverty adopted by the Federal government in 1965. The standard has two parts: first, a set of income thresholds (poverty lines) for families of different sizes, and, second, a definition of family income to be compared to the threshold. The 1965 standard is what economists call absolute and objective, objective in the sense that poverty is measured by a family’s income rather than the member’s thoughts or feelings and absolute in the sense that whether one family is poor is independent of how other families are doing. As a result, given the official standard, every family in the United States could have been living in poverty in 2001 or 2002.

The official U.S. standard is one of many ways poverty could be measured (Ruggles 1990). Poverty could be measured in relation to median family income. Many international organizations take a family to be poor if and only if the family’s income is less than one half the country’s median income: this is an objective but relative (comparative) measure of poverty. According to this measure, not every family in the United States could be living in poverty. Poverty could also be measured by a family’s consumption rather than income or measured by what members believe they need to keep up with their neighbors; this is a subjective measure. A person can feel needy or think she is less well off than others no matter what her income.

Given a variety of standards to choose from, anyone who wishes to study how a risk within a population varies with poverty has to decide which of the standards to employ. Moreover, any estimate of the error in a poverty statistic relies on the choice of standard, since any error in a measurement of the number of poor is the difference between the number taken to be and actually poor, and the number who are actually poor depends on the standard of poverty one has chosen to employ. Which standard is best depends on one’s interest. In other words, to distinguish between a family’s actual and apparent economic status, an economist has to choose between the different standards of poverty, but since she must base her choice on her interests, the actual status of the family will be based on that interest as well.

Whether a family is poor depends on one’s interests in surveying the population and not simply on facts about the family. Race, on my view, is similar in this respect. We are used to thinking of race as an intrinsic property of members of a population, but there is no more reason to think of a racial variable this way than there is a poverty variable. Neither variable ranges over a set of simple demographic facts, and no way of measuring the variable is intrinsically better than any other.

For the sciences to assign members of a population to a category like “poor” or “White” whenever they say they are is a mistake if how they identify themselves does not have much effect on the risk the variation of which is to be
described or explained. Whether a subjective or objective measure of poverty or race measures the actual value of the variable depends on the risk and can vary from one risk to another. An economist cannot capture something as complex as poverty with a single measure or standard of poverty, but an epidemiologist cannot capture something as complex as race with a single way of assigning members of a population to one of the five OMB categories.

Race, like poverty, has a number of dimensions and one way of assigning individuals to the category can be a better measure of one dimension and a worse measure of another. Discrimination is one dimension of race and ancestry another one. Other-reported race might be the best measure of an individual’s risk of discrimination but mother’s race a better measure of his ancestry. As a result, other-reports can be the best measure of race in a study of how access to health care within a population varies with race, but mother’s race a better measure in a study of how race affects the risk of a recessive genetic disorder.

7. Race and Ancestry

According to some genetic epidemiologists, self-reported race is the optimal way to categorize humans in the United States by race for genetic research (Risch, Burchard, Ziv, and Tang 2002); they assume that self-reported race is a reasonable proxy for ancestry and that differences in ancestry describe or explain differences in the risk of heritable diseases or differences in how drugs are metabolized. However, their assumption rests on a sample for which self-reports, observer reports and parent-based measures are most likely to identify the same subpopulations, but, with respect to an increasing number of Americans, the different ways of identifying race identify different subpopulations. To the extent that an increasing number of Americans are of mixed race, recent immigrants, or vary the race they assign themselves from context to context, self-reported race becomes a poor surrogate for ancestry and a poor measure of the actual race of members of the U.S. population in relation to any medically relevant genetic differences between them.

Even if self-reported race were a reasonable proxy for ancestry within the U.S. population, self-reported race would not be an optimal way to categorize the population in order to study differences in mortality and morbidity since few diseases are inherited or primarily the result of a gene. Moreover, even with inherited diseases, the risk of the disease varies not only with the genes but also with the environment or with heritable changes that regulate gene activity in the absence of differences in a DNA sequence, and some of these factors vary more with other-reported than self-reported or mother’s race (Foley et al. 2009).

Let E be the variation between racial groups in a social or biomedical trait T within a population and A the variation in the trait within racial groups. The greater the ratio of E to A (i.e., the larger E/A), the better race describes the variation in T. As a result, given an interest in describing or explaining variations in T within a population, which of the four ways to assign race to members is best
is a matter of which results in the largest ratio. If the ratio is larger with other-reported than self-reported race, then other-reported race is a truer measure of race in relation to an interest in T than self-reported race. While the ratio might be greater with other-reported than self-reported race for one T, the ratio might be less for another.

If the difference in rate of death due to sickle-cell disease between a group whose mother’s race is Black and a group whose mother’s race is White is greater than the difference between a group of self-reported Blacks and a group of self-reported Whites, then mother’s race rather than self-reported race should be taken to be the actual race of members, given an interest in describing or explaining differences in the rate of sickle-cell deaths within the population.

8. Objections Reconsidered

By allowing the actual race of members of a population to vary, a social or biomedical scientist who uses race to study risk can avoid the objections to the use of race I mentioned at the beginning of this article. According to the first objection, since there is no consensus with respect to how to assign race to members of the U.S. population, “race” can have no scientific meaning and reports of racial differences in the risk of death and disease or poverty and unemployment no scientific significance. However, if the actual value of a racial variable is allowed to vary with interest, and race is taken to be more like poverty than place of birth, then there can be a consensus with respect to how to assign race to members, and a report of racial differences in the risk of a particular disease or disease-related death can have a clear and agreed-upon meaning. The report can be understood to mean, for example, that the rate of the disease is greater among the members who self-report one race, for example, Black, than among those who self-report another, for example, White. A report of racial differences in the rate of a different disease, on the other hand, can be understood to mean that the rate is greater among members whose mothers are one race than among members whose mother’s race is another. Though in relation to each disease, a member has an actual race, her actual race can be different from one disease to another.

According to the second objection, since racial groups are not natural biological kinds or genetically distinct populations, the practice of grouping by race rests on a false belief that there is a genetic property (a race gene or gene cluster) that distinguishes the races. However, if members of a population classify one another by race and some racial groups are exposed to more health risks than others, no genetic property has to be shared by every member of the group in order for race or racial categories to have a meaning or be useful in biomedical research. The groups can be real even if the members do not differ in the way that many of the members assume they do.7

According to the third objection, the use of race encourages racialist thinking or racial stereotypes. However, if the race of a member is understood to depend on the risk, and to vary with the risk, the use of race does not reinforce but opposes
racial stereotypes and racialist thinking, since to understand that a person’s actual race can vary is to recognize that race is not an intrinsic property of persons but a relation between a person and a risk or an interest.

9. Conclusion

Race, on my view, should continue to be used as a variable in biomedical research, as long as the risk of morbidity and mortality within the population varies with self-reported, other-reported or mother’s race, but the method currently used to assign members of a population to a racial category (always self-reports) should be changed; assignments of race should be allowed to vary. The change I recommend would make research in the biomedical sciences somewhat more difficult, for racial differences in the risk of one trait might not be comparable to differences in the risk of another, but the change would make a finding of difference more salient. While an individual might have one race in a study of one biomedical trait and a different race in another, a correlation between race and a biomedical trait would have more descriptive or explanatory significance if race were treated as a variable rather than fixed characteristic of persons.

Notes

1 On the nature of race, see, for example, Andreasen (2005), Mallon (2004), Haslanger (2000), and Root (2000). On the bearing of race on issues of justice, see for example, Boxill (1992) and McGary (1999).

2 See Kington and Nickens (2001).

3 There is some controversy over the best term to use to describe differences between social groups in health. Some researchers use the term “differences,” others use “disparities,” and others “inequalities” or “inequities.” The last terms are often frowned upon because they seem to invoke considerations of justice or fairness and, on the view of many, research in the social and biomedical sciences should be free of judgment of moral and political value (see Root 1993 for a discussion of this view). As some researchers use the term “disparities,” not all differences are disparities but only those differences that could be affected by changes in public policies, and, as some use “inequalities,” inequalities are disparities between advantaged and disadvantaged social groups (between groups that have and groups that have not experienced unfair treatment or discrimination). For a discussion of the different terms used to describe health differences see Braverman (2006).

4 According to the National Center for Health Statistics (NCHS), while 14.6 percent of Whites in the United States (regardless of health insurance) had no usual source of health care in 2006, 17.9 percent of Black had no usual source of care. As long as members of the U.S. population who are taken to be Black have different access to health care than those taken to be White, epidemiologists need to employ race as an analytic variable in their studies of disparities in morbidity and mortality.

5 Often the researcher does not collect any data but relies on data collected by the government. In such a case, the decision to which category to assign a member of the population is made before she begins her study; however, in the government data sets she relies on, race has been assigned to members of the population in one of the same four ways, viz. self-report, other-report, mother’s race, or race on a death certificate; in most cases by self-report.
According to Waters (2000), foreign-born are less likely than native-born Americans to assign themselves the race others do, since native-born Americans are more accustomed to using the OMB categories to classify people, while foreign-born Americans are more accustomed to using ethnicity or nationality to identify themselves.

Some philosophers take this to mean that race is not real but racism is. However, within the context of biomedical research, whether race is real is not a matter of whether members of different racial groups differ more genetically than members of the same group but whether they differ more in their exposure to a health risk or in their rate of morbidity and mortality.

References


