The Number of Black Widows in the National Academy of Sciences

Michael Root†

Studies in the social and biomedical sciences of racial differences in socioeconomic status or health within a population view the race of members as fixed and look for a difference in the frequency of a trait like average income or disease risk between racial subgroups. But, as I explain in this paper, there are good reasons to allow the race of members to vary with the trait whose variation within the population is to be described or explained. According to such a view of race, racial categories are more scientifically significant if membership in the categories is allowed to vary with differences in scientific interest rather than held constant across a variety of interests.

1. Introducton. The National Academy of Sciences, according to a recent report in the *Chronicle of Higher Education*, has 1,922 members; 160 are women and four members are believed to be black (Brainard 2003). The *Chronicle*'s report is noteworthy in at least two respects: first, the small number of women, and, second, the uncertain number of blacks. While the *Chronicle* reports the number of members who are believed to be black, the number of actual blacks in the Academy is left open; we are left to wonder how many members are actually black and how the actual race of each of the 1,922 members is to be decided?

Race is an important descriptive and analytic category in the social and biomedical sciences. Sociologists study differences in the United States between the races in median income or average test scores, and epidemiologists study differences between black and white infants in average birth weight or the risk of diabetes. Race is often used, in other words, as a variable in descriptions or explanations of differences between mem-

†To contact the author, please write to: Department of Philosophy, University of Minnesota, 831 Walter Heller Hall, 271 19th Avenue South, Minneapolis, MN 55455-0310; e-mail; mdroot@umn.edu.

bers of the population in an important measure of prosperity or health, and public policy is often based on these descriptions or explanations.[1]

Most studies assume that race is a fixed characteristic and that the size of racial populations vary little with how a race is assigned to individuals or how a black member is distinguished from a white; they take the racial identification of members as fixed and try to describe or explain a difference in the frequency of a socially or medically significant trait between individuals identified as racially different. For many individuals in the U.S., however, race is not a fixed characteristic but varies depending on how race is assigned, and, in particular, whether the assignment is based on the race the individual reports herself to be, the race she is most often assigned by others or her mother's race.

As a result, statistical measures of racial disparity can be an artifact of the way race is counted rather than a description of a real difference between races in a socially or medically significant trait, and the sciences need to consider how a race should be assigned to a member of a population when describing or explaining differences between racial groups in the trait. Given the complexities of racial identity in the U.S., there is no single best way, I argue, for individuals to be racially categorized. How best to assign race depends on our particular interest or on the character of the trait whose variation within the population the science is trying to describe or explain.

In this respect, race is different from many other variables employed in the sciences, for the actual value of variables like household income and SAT scores or birth weight and diabetes do not vary with interest as the actual value of the race variable does. According to some demographers, the usefulness of race or ethnicity as a category in the sciences depends on a uniform assessment of racial or ethnic status across federal and state data systems (Williams 2000). However, a uniform assessment, I argue, reduces the descriptive or explanatory power of race as a category in the social and biomedical sciences. As a result, we should not expect all data sets to agree in their assignments of race or different sets of racial data to be comparable.

---

1. By using the number of members in a population with a particular social or biomedical trait T, e.g., diabetes or unemployment, as the numerator and the number of members of a given race as a denominator, the social or biomedical sciences are able to describe the difference in risk of T between the different races. Some economic and health disparities between members of a population are often more adequately described or explained by differences in age, socioeconomic status, education, ethnic identification or some number of other social or psychological variables than they are by race, but many of them persist when variables other than race are controlled for. As a result, race remains an important variable in descriptions and explanations in the sciences.

2. Race and Marital Status. By matching social security records and marital information collected from a variety of surveys, demographers have discovered that survey-reported marital status is inaccurate, for divorced survivors of an ex-spouse often report themselves as widowed, when (given the rules employed by the Social Security Administration for assigning marital status) only current spouses at the time of death count as widows or widowers (Weaver 2000). Demographers are able to distinguish between the members of a population who are widowed and the members who think themselves to be but are not, to the extent that the Social Security Administration is taken to be the arbiter of marital status. Moreover, they are able to use standard statistical methods to estimate the measurement error in studies using self-reported marital status to explain differences within a population in a particular trait, since they are able to measure the difference between the reported values of the marital variable and the actual or underlying values.

But why should a demographer accept the designation of marital status on a social security record as the real marital status of a member of a population? The Social Security Administration has an interest in seeing that once a man is deceased, his social security income is not divided between two beneficiaries, but demographers do not have that interest, and their studies might be better served by counting a surviving ex-wife or domestic partner as a widow.

By matching race on birth records and the self-reports of race collected as part of the 2000 Census, demographers could discover that the self-reports of race, like the self-reports of widowhood, are inaccurate, if race on a birth record is taken to be the real race of members of the population.[2] Moreover, they could use standard statistical methods to estimate the measurement error in studies using self-reported race to explain differences within a population in a particular trait, since they could measure the difference between the reported values of the race variable and the actual or underlying values.

But why should demographers treat a person's race on his birth record as his actual or underlying race? Were they to take self-reports rather than birth records to be the actual measure of race, demographers could use standard statistical methods to estimate the measurement error in studies using race on birth records to explain differences within a population in traits like income or unemployment, since they could measure the difference between race as measured by each of the two criteria.

The authors of one recent study of racial differences in birth weight comment that measurements of race, like measurements of other social

---

2. A person's race on his birth record is, as a rule, the race his mother took herself to be at the time of his birth.

science variables, can contain errors that attenuate a correlation between one variable (in this case race) and another, viz. birth weight (van Den Oord and Rowe 2000). But the comment is misleading, for race is different from many other variables employed in the sciences, since, with race, all we seem to have is imputed values. That is, there is no apparent reason to take a person's self-reported rather than her observer-reported race or parents' race to be her actual or underlying race, no reason to say that a person's actual race is her race on her birth certificate rather than the race she reports herself to be.

Within the social or biomedical sciences, one customary way to count blacks or widows is intrinsically no better than another but can be better in relation to a variable whose variation within a population the science wishes to describe or explain. Epidemiologists who wish to describe or explain how hypertension varies between wives and widows over 40 have to decide whether a woman divorced from a deceased husband is a widow or not. Should they believe that women with deceased husbands and women with deceased ex-husbands are similarly exposed to the causes of hypertension, they have a reason to place the women in the same category, even if the Social Security Administration does not. Should they believe that their exposure is different, they have reason not to, even if the women all place themselves in the same category. Moreover, epidemiologists could have a reason to count them all as widows with respect to one disease but not with respect to another if they believe that their exposure to the causes of one disease is the same whether they are divorced from the deceased or not, but their exposure to the other is different.

3. The Trouble with Self-Reports. The social and biomedical sciences increasingly rely on self-reports when assigning race to members of a population (Friedman et al. 2000, 1715). Many epidemiologists call self-reported race "the gold standard" when studying how rates of morbidity or mortality vary with race and recommend that, whenever possible, self-reports be used to assign race when collecting racial data (Kaufman and Cooper 2001; Jones 2001). Assigning an individual the race she assigns herself is often the easiest or most respectful way to assign her a race. 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 social or biomedical scientists to identify the race of their subjects might not give race as much descriptive or explanatory power as a less easy or respectful way.

Other-reported race better explains differences in a social or biomedical trait within a population to the extent that the differences are primarily due to a member's exposure to racial discrimination, since his exposure

to racial discrimination is not based on the race he assigns himself but on the one he is assigned by others. As a result, whenever differences between members of a population in a biomedical trait are likely to be due to racial discrimination, self-reported race should be employed as a gold standard only if there is good evidence that self-reported race is a proxy for other-reported race.

A workshop at the National Academy of Sciences concluded in 1996 that research is needed to assess the data compatibility between racial identification done by self-reports and the reports of others (Edmonston et al. 1996). Studies conducted since suggest that the self-identified race of many members of the U.S. population is fluid rather than fixed, even if their other-identified race is fixed, and that, in the case of recent immigrants and the children of parents of different races, self-reported and other-reported race are often different (Harris and Sim 2002).

Multiracial individuals, according to the National Longitudinal Study of Adolescent Health, frequently assign themselves a race different from the one they are thought to be by interviewers; for example, only 67% of the children of black-white unions who identified themselves as white were identified as white by interviewers, while 95% of those who identified themselves as black were identified by interviewers as black (Harris 2000). If 33% of the multiracial children who report being white are exposed to no less discrimination than children who report being black, then, in these cases at least, other-reported race is a better marker of exposure to racial discrimination than self-reported race and, as a result, a better predictor a child's risk of morbidity, to the extent that rates of morbidity among children vary with exposure to discrimination.

In 2002, six states and the District of Columbia piloted a "Reactions to Race Module" on the Behavioral Risk Factor Surveillance System (BRFSS) that includes the question, "How do other people usually classify you in this country?" followed by the OMB race categories and the category "Hispanic and Latino" (CDC 2002). Since the BRFSS also asks the OMB ethnicity and race questions, social scientists should be able to compare the responses and measure the degree of correlation between imputed other-reported and self-reported race. Should the correlation be weak and the correlation between imputed other-reported race and other-reported race be strong, they would have evidence that self-reported race is not a very good proxy for the race individuals are assigned by others.

4. The Fluidity of Race. Prior to 1960, the U.S. Census enumerated race based on phenotype (most census takers inferred a respondent's race from her skin color and other bodily features); in 1960, the practice of self-definition began, and all members of a household were counted as black if the head of household reported being black. In the 1960 census, many

Hispanics who had been counted black in 1956 did not report being black, especially if they arrived in the U.S. from a Spanish-speaking country in which they were not as counted black. As a result, in the case of foreign-born Hispanics, at least, self-reported and observer-reported race are often different.

Hispanics, like mixed-race Americans, seem to have a variety of context-specific racial self-identities, but, despite their self-identities, if many are consistently identified as black by others, then a social scientist who is interested in how access to housing, education, mortgage lending, health-care services or employment opportunities vary with race has a reason to take other-reports rather than self-reports as the best measure of the person's race and to stratify the population by the race the members assign to one another rather than those they assign themselves.

Other-reported race is also fluid. An individual can be identified as one race by me and a different race by you. When describing or explaining how a trait varies with a difference in other-reported race, a social scientist or epidemiologist should consider whose other-reports are most likely to affect the trait. Where the trait is the risk of a traffic stop, for example, reports by police officers of a motorist's race should be the gold standard in assigning race, and where the trait is the risk of invidious discrimination in employment, reports by employers of a worker's race should matter most, when the social scientist assigns race to the members of that population.

In the 2000 Census, 6.8 million people, or 2.4% of the total U.S. population, reported having two or more races. According to a rule adopted by the Bureau of Census on the aggregation and allocation of multiple race responses for use in civil rights monitoring and enforcement, responses that include one minority race and "white" are allocated to the minority race. Why the minority race? Because, in the context of civil rights, what matters most is not whether a person sees herself as white but whether others see her as a minority. If a person who sees herself as both "white" and a member of a minority race is treated as a minority by members of the majority who practice racial discrimination, then the agencies that monitor and enforce the civil rights laws have a reason to identify as black anyone who identifies herself as both black and white.

As more questionnaires or surveys allow respondents to report more than one race, or as the mixed race movement in the U.S. grows, more people will identify themselves as more than one race, and the less reasonable it will be for the social or biomedical sciences to treat self-reported race as a marker or proxy for a person's exposure to racial discrimination. As a result, the reason behind the current rule for aggregating and allocating multiple responses to the race question in the U.S. Census is also a reason for scientists to favor other-reported over self-reported race in

BLACK WIDOWS IN THE NATIONAL ACADEMY OF SCIENCES   1203

assigning race to individuals; instead of asking members of a population what race they assign themselves, a sociologist or epidemiologist should ask them what race other people usually take them to be, whenever there is reason to suspect that differences between members in a trait is primarily due to racial discrimination.

When the U.S. Bureau of Census classifies as black respondents who self identify as white and black and adds them to the count of blacks rather than whites, they are not failing to count the actual number of blacks or whites in the population; for the census, like any other data set, captures an individual's actual or underlying race relative to a particular purpose, and, in the case of the census, one major purpose is to monitor civil rights (Harris and Sim 2002, 625). Whether a data set in the social or biomedical sciences captures the actual or underlying race of members of a population also depends on the particular purpose for which the data are collected. Usually, in the sciences, the purpose is to describe or explain a variation in a socially or medically significant trait within a population. Relative to describing or explaining the variation in one trait, a member might be identified best as a member of one race, but relative to describing or explaining the variation in another, she might be identified best as a member of another.

When self-reported and other-reported race differ within a population of high school students, for example, self-reported race might best explain each student's choice of associates, while his other-reported race might best explain why the other students have chosen or not chosen to associate with him. While self-reported race might best explain a difference between the students in graduation rates, other-reported race might best explain a difference between them in median family income (Fordham 1986). As a result, from the perspective of a science, no less then the perspective of the population the science has chosen to study, race should be understood as a fluid rather than fixed characteristic of persons, and a person's actual race should be allowed to vary with the socioeconomic or biomedical trait the scientist wishes to study.

5. Racial Explanations. Differences between racial or ethnic groups in the risk of death or a disease 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 (typically by self-report), 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; the races, no matter how we define them, are not natural kinds, and the members do not share any core properties; 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 the rule used to assign race to members of the population, and the best rule is the one by which the difference between inter-group and intra-group differences in the trait is greatest.

If D is an inherited disease, then differences in (a) better explain differences between racial groups within the population in the incidence of D than differences in (c) or (d) do if membership in the group is identified on the basis of parent-based measures of membership rather than self-reported membership, whenever the two pick out different members, since parent-based measures are better proxies for genetic ancestry, and (inherited) genetic differences between members are due to differences between them in ancestry rather than differences in how they identify themselves.

In the case of recently admixed populations, genetic ancestry varies significantly among members with the same self-reported race, and, as a result, self-reported race is not a good a way to determine the race of members of these populations if the trait whose variation is to be described or explained is genetic. That is, many members of the U.S. population who identify as one race, e.g., black, are of mixed ancestry with origins in the indigenous people of more than one continent and, to the extent that the distribution of (a) varies significantly between the different indigenous populations, the distribution of (a) will vary significantly within a group with mixed ancestry whose members self-identify as black.

According to some genetic epidemiologists, self-reported race is the optimal way to categorize humans in the U.S. by race for biomedical research (Risch et al. 2002); they assume that self-reported race is a reasonable proxy for ancestry, but their assumption rests on a sample for which self-reports, observer reports and parent-based measures are most likely to identify the same subpopulations, and, as I argue in this paper, 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 or vary the race they assign themselves from context to context, self-reported race becomes a poor surrogate for ancestry.

Different ancestral populations do differ in the genes that encode drug-metabolizing enzymes or in the genes responsible for disease (McLeod 2001). Race can serve as a proxy for these populations if but only if the average number of ancestors in each of two populations that differ in the distribution of these genes differs between different racial groups. In such cases, differences in genes can explain why different racial groups differ in risk for an inherited disease or metabolize a drug differently than members of another.

While differences in genes can explain differences in the risk of an inherited disease between the races, differences in race do not explain the differences in the risk of the disease in the population unless differences in race explain differences in (a). Even if race describes differences in the population genetic structure of variable drug response or disease, race does not explain the differences in structure unless, as a result of racial discrimination or opposition to racial mixing, one racial group, for a time, is reproductively isolated from another.[3] In such a case, while ancestry may be a proxy for differences in (a), other or self-reported race is a proxy for barriers to racial mixing.

However, if D is an infectious disease, then differences in (b), (c), or (d) are more likely to explain differences in D between racial groups if membership is based on other-reports than self-reports, whenever the two pick out different members. Since individuals who are exposed to racial discrimination are more likely to live or work in areas in which contact with the disease-causing toxins, viruses or bacteria is high than individuals who are not, and other-reported race correlates better with exposure to racial discrimination than other ways of assigning race, epidemiologists, in describing or explaining racial disparities in D, have a reason to choose other-reported race as the actual or underlying race of the members of the population. Differences in race can explain (as well as describe) differences in disease risk or drug response due to differences in (b), (c), or (d) (unlike differences in risk or response due to (a)), since differences in race explain differences in employment, schooling, housing, and healthcare, and these explain differences in exposure to the agents responsible for the differences in risk or response.

While blacks are more likely to have high blood pressure than whites because they are black, they are not more likely to have sickle cell disease because of their race, even though blacks in the U.S. are 150 times more likely to have that disease than whites are (Tapper 1999). Sickling is an inherited hemoglobin disorder and the result of a single recessive gene; children who inherit the gene from only one parent have the sickle cell trait, and those who inherit it from both are likely to have the disease. That is, individuals who are heterozygous for the gene have the trait, while only individuals who are homozygous have the disease.

The frequencies of the allele for the trait are high in populations that

3. Should racial discrimination prevent individuals who happen to differ at some gene loci from breeding with one another, race as (as a social status) might be said to be a (social) cause of differences between two racial groups in gene frequencies in the U.S. today and, in particular, in genes responsible for D, and racial discrimination, in such a case, might be said to be (causally) responsible for the genes that members of each group in the U.S. have today.

have been continuously exposed to malaria because being heterozygous for the trait confers resistance to the malarial parasite in regions like southern India and central Africa. A larger proportion of blacks than whites in the United States have the sickle cell trait due to migration rather than the race; during the Atlantic slave trade many people from malarial Africa were kidnapped and taken to America and only a few whites and Asians migrated to the United States from malarial regions of Europe or India. As a result, within our borders, sickle cell anemia looks like a black disease. To the extent that differences in race were a cause of differences in migration, e.g., an African would not have been kidnapped, sold into slavery and taken to the Americas had she been white rather than black, race enters into the explanation of why blacks are 150 times more likely to have sickle cell disease in the U.S. than whites (Root 2003, 1176–1177).

Race can be a social cause of a disease but not a genetic cause, since there is discrimination for race but no gene for race. While race, in the U.S., is often a cause of differences in (b)–(d), race is not a cause of differences in (a) unless differences in (a) are due to racial discrimination. Race can serve to pick out people in the U.S. who are at high risk for a genetic disease like sickle cell, but race can only explain why the risk is high if race can explain why groups at high risk are more likely than groups at low risk to have entered the U.S. from regions in which the incidence of the sickle-cell gene is high.

6. Conclusion. On the view of social variables like race and marital status that I have offered in this paper, the number of black widows in the National Academy of Science is uncertain, for the best way to draw the distinction between members who actually are widowed or black and those who are only thought to be depends on one's interest in stratifying that population by race or marital status. I draw a similar conclusion with respect to the use of race as a descriptive or analytic category in the social or biomedical sciences; the actual or underlying race of members of a population studied by a science depends on the trait whose variation within the population the science is interested in describing or explaining. According to conventional wisdom, a person's race is fixed, and each time she is assigned a race, the race she is assigned should be the same. But a relative conception of race can be better than a fixed one when trying to describe or explain differences in mortality or morbidity within a population, for while the race of a person's mother can increase her risk for a genetic disease, the race she is most often taken to be by others can increase her risk for an environmental one.

REFERENCES

Brainard, Jeffrey (2003), "National Academy of Sciences Elects Record Number of Women", *Chronicle of Higher Education*, "Today's News," April 30 (http://chronicle.com/prm/daily/2003/04/2003043003n.htm).

Centers for Disease Control and Prevention (CDC) (2002), *Behavioral Risk Factor Surveillance System Survey Questionnaire*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Edmonston, Barry, Joshua Goldstein, and Juanita Tamayo, eds. (1996), *Spotlight on Heterogeneity: The Federal Standards for Racial and Ethnic Classification*. Washington, DC: National Academy Press.

Fordham, Signithia, and John Ogbu (1986), "'Black Students' School Success: Coping with the Burden of Acting White", *Urban Review* 18: 176–206.

Friedman, Daniel, Bruce Cohen, Abigail Averbach, and Jennifer Norton (2000), "Race/Ethnicity and OMB Directive 15: Implication for State Public Health Practice", *American Journal of Public Health* 90: 1714–1719.

Harris, David R. (2000), "Demography's Race Problem", paper presented at the 2000 meeting of the Population Association of America (http://www.nichd.nih.gov/about/cpr/dbs/sp/harris.htm).

Harris, David R., and Jeremiah Joseph Sim (2002), "Who Is Multiracial? Assessing the Complexity of Lived Race", *American Sociological Review* 67: 614–627.

Jones, Camara Phyllis (2001), "Invited Commentary: 'Race', Racism and the Practice of Epidemiology", *American Journal of Epidemiology* 154: 299–304.

Kaufman, Jay S., and Richard S. Cooper (2001), "Commentary: Considerations for Use of Racial/Ethnic Classification in Etiological Research", *American Journal of Epidemiology* 154: 291–298.

McLeod, Howard L. (2001), "Pharmacogenics: More than Skin Deep", *Nature Genetics* 29: 247–248.

Risch, Neil, Esteban Burchard, Elad Ziv, and Hua Tang (2002), "Categorization of Humans in Biomedical Research: Genes, Race and Disease", *Genome Biology* 3: 1–13.

Root, Michael (2003), "The Use of Race in Medicine as a Proxy for Genetic Differences", *Philosophy of Science* 70: 1173–1183.

Tapper, Melbourne (1999), *In the Blood: Sickle-Cell Anemia and the Politics of Race*. Philadelphia: University of Pennsylvania Press.

van Den Oord, Edwin J. C. G., and David C. Rowe (2000), "Racial Differences in Birth Health Risk: A Quantitative Approach", *Demography* 37: 285–298.

Weaver, David A. (2000), "The Accuracy of Survey-Reported Marital Status: Evidence from Survey Records", *Demography* 37: 395–399.

Williams, David R. (2000), "Race/Ethnicity and the 2000 Census: Recommendation for African American and Other Black Populations in the United States", *American Journal of Public Health* 90: 1728–1730.