LIMITATIONS IN THE USE OF RACE IN THE STUDY OF DISEASE CAUSATION

Richard S. Cooper, MD, and Vincent L. Freeman, MD
Maywood and Hines, IL

Tremendous variation exists in the rates of many chronic diseases across racial groups. However, serious technical and conceptual limitations hamper the ability of racial comparisons to illuminate the causative pathways. First, race is confounded by social class, which is complex, and like other confounders of race, may not be measured with equal validity across racial groups. Second, statistical "adjustments" for race effects can be misleading since residual confounding may be misconstrued as a genetic effect. Third, the biologic concept of race tempts us to ignore the context dependency of genetic expression. When trying to detect genetic effects, both the environmental and genetic contributions must be measured and potential gene-environment interactions accounted for. Unfortunately, this process is beyond our current technical capabilities.

To move forward on the problem of prostate cancer and other diseases distinguished by marked ethnic differentials, investigators need a more comprehensive understanding of the factors that mediate the apparent effect of race combined with valid measures of those factors, as well as novel strategies that can help overcome the technical and interpretive limitations of statistical adjustment. Finally, the "grand" theories of race-based genetic susceptibility must be replaced with rigorous criteria to determine when a trait can be ascribed to some genetic origin. (*J Natl Med Assoc.* 1999;91:379-383.)

Key words: race ◆ disease ◆ etiology ◆ environment ◆ genetics

Tremendous variation exists in the rates of many chronic diseases across racial groups.[1] Investigators have made use of these differences for several purposes, with the most fundamental being for the understanding of etiologic relationships.[2-4] The concept of race as a biologic model providing genetic explanations for disease has made this approach particularly appealing to some for the study of genetic effects in prostate cancer.[5] While race may be useful in defining where the problem is, using it to test etiologic hypotheses is inherently problematic. Race, or ethnicity, is a proxy for determinants of disease; it alone does not represent a primary pathway or mechanism. As a result, ethnic comparisons to elucidate disease causes face major design and analytic obstacles, namely, the presence and measurement of confounders, the reasonableness and adequacy of statistical "adjustments" for race effects, interpreting the meaning of residual confounding, and accounting for interactions.

To provide examples of these technical and conceptual limitations and thus shed light on the limitations of the biologic concept of race in etiologic investigation, this article focuses on two essential areas used

From the Department of Preventive Medicine and Epidemiology, Loyola University Stritch School of Medicine, Maywood, and the Department of Veterans Affairs, Midwest Center for Health Services and Policy Research, Department of Veterans Affairs, Edward Hines, Jr Hospital, Hines, Illinois. Presented at the Leadership Conference on Prostate Cancer in the African-American Community, November 20-22, 1997, Houston, TX. Requests for reprints should be addressed to Dr R. Cooper, Dept of Preventive Medicine, Loyola University Stritch School of Medicine, Maywood, IL 60153.

Table 1. Mortality Rates* for Selected Cancers in Whites, Blacks, and Hispanics, 1990†

	Men			Women		
Type of Cancer	White‡	Black	Hispanic	White	Black	Hispanic
Lung	83.5	116.2	56.8	39.9	34.2	20.1
Colorectal	23.3	32.4	22.8	17.5	22.9	15.0
Prostate	31.2	63.9	26.0	—	—	—
Breast	—	—	—	29.4	33.4	25.0

*Per 100,000 persons.
†From reference 1.
‡Non-Hispanic whites.

Table 2. Education, Income, and Financial Assets Among US Blacks and Whites*

	Income		Financial Assets	
Education	White†	Black	White	Black
Some high school	$11,554	$8724	$1100	$0
High school	17,328	11,534	3287	0
Some college	27,594	21,076	5500	0
College degree	35,068	28,080	17,300	5
Postgraduate	40,569	31,340	23,200	78

*Source: Oliver & Shapiro, *Black Wealth/White Wealth*, 1995.
†Non-Hispanic whites.

to measure racial effects: social class and genetics. We will identify issues that need to be resolved to move forward not only on the problem of prostate cancer but also of other chronic diseases characterized by marked ethnic differentials.

Large differences in cancer disease rates by ethnicity have been well documented.[1] Table 1 shows the ratio of cancer death rates, which are considerably higher among blacks across virtually all tumor sites and somewhat lower for Hispanics. A common alternative hypothesis proposed to explain such racial/ethnic variation in disease is that the observed differences are not due to race per se, but social class, often measured as socioeconomic status (SES).[6] Thus, race is "confounded" by social class.

Confounding is a mixing of effects.[7] Specifically, the estimate of the effect of the exposure of interest is distorted because it is mixed with the effect of a third factor. For the third factor to be a confounder, it must be a correlate of the exposure and a correlate of the outcome (although not necessarily in a causal fashion). Thus, in this context, "race" is the exposure of interest, "cancer death rate" is the outcome of interest, and "social class" is the third factor, which is a correlate of both exposure (race) and outcome (cancer death rate). However, as Figure 1A suggests, these data are not entirely consistent by our general measures of social class: on the one hand, we see a relationship between blacks and whites, and lower SES is associated with higher mortality; however, that is not true for Mexican Americans. One of the problems is that SES, like race, is not a direct measure but a proxy of certain lifestyles or patterns. Thus, if dietary patterns are suspected to be related to cancer risk, we might see the relationship between cancer mortality and race if SES is replaced with the causal variable diet as measured by intake of fruits and vegetables (Figure 1B).

CROSS-ETHNIC COMPARISONS

Two major assumptions are involved when making cross-ethnic comparisons. First, it is assumed that the potential operative variables of interest can be measured with comparable validity across ethnic

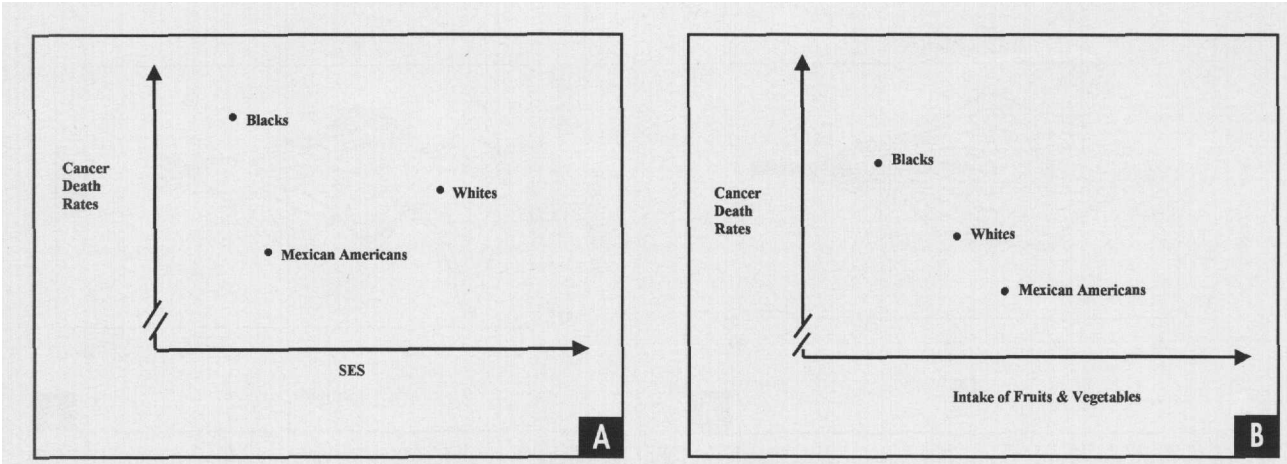

Figure 1. Relative cancer death rates in US ethnic populations in relation to socioeconomic status (SES) (A) and diet (intake of fruits and vegetables) (B).

groups, and second, the variable measured has a similar relationship within each group. As a result, it then would not be necessary to take into account interactions between race and other explanatory variables. An interaction can be thought of as the condition where the presence of one factor alters the relative effect of another.

Unfortunately, the aforementioned assumptions about measurement validity and the absence of interactions are usually quite difficult to satisfy. For example, income and education are believed to be the primary determinants of social class.[8] However, other conditions or attributes, such as accumulated wealth, may, in fact, be how income and education mediate their effects, which also may vary across racial groups. Table 2 shows that while among whites, financial assets rose significantly with education, this was not the case among blacks, where there is virtually no correlation between education and financial assets. Ultimately, it is financial assets that determine one's ability to buy a house, send children to college, etc. Figure 2 provides another example of what has been referred to as an "interaction." Socioeconomic status, as measured by income, is plotted against both death rates and age, creating a diagram of the relative risk of dying at different SES levels. These data show that by middle age, SES, as measured by income, is a major predictor of the risk of dying. However, by retirement age, the measure of income is a predictor neither of risk or social class. Hence, this effect of income varies across age groups, ie, exhibits an interaction with age.

In summary, ethnic comparisons are confounded by SES, a measure of social class. Socioeconomic status is multifaceted, poorly defined, and difficult to measure. Furthermore, it often is approached in categories, and there is substantial potential for interactions, ie, variation of its effect across those categories. What are the implications for the use of the biologic concept of race posed by the generic technical problems that these examples illustrate? A discussion of the standard analytic approaches used in epidemiology used to solve these problems should provide some insight into this question.

EPIDEMIOLOGIC CONSIDERATIONS

Statistical adjustments are used to account for the effects of known confounders when comparing groups. After this "adjustment," the difference that remains between the groups being compared is referred to as "residual confounding."[9] In epidemiology, when comparing blacks and whites, what is often done is to adjust for the effect of SES and other confounders, and interpret any residual difference as a "race effect." The temptation is to infer that what is left over after environmental adjustments must be genetic.

For example, one might make the argument that the racial difference observed in prostate cancer disease rates after adjusting for SES is an intrinsic attribute of blacks compared with whites. The logic of this argument, however, is fatally flawed because there are many attributes bound up in the social development of race that play out in our culture and simply cannot be removed by "adjustment." And yet, the medical literature is filled to overflowing with analyses that conclude with these arguments.[10-15]

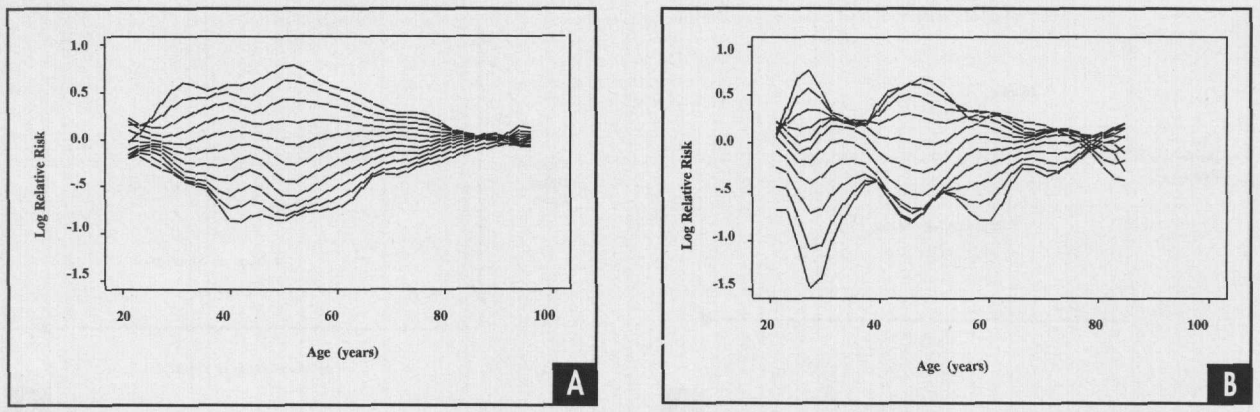

Figure 2. Socioeconomic status (income/poverty level) and the risk of death at different ages among white (A) and black (B) men in the United States: National Health Interview Survey, 1986-1990.

Furthermore, for a given race-disease relationship, all of the relevant environmental factors are never known. Therefore, to the extent to which these effects can be inferred in this fashion, statistical adjustments inherently overestimate the presence of genetic effects. Consequently, for the basis of ethnic comparisons, the method of statistical adjustment, which is the fundamental approach used in epidemiology and in medical science generally, is inadequate and usually unreasonable

GENE-ENVIRONMENT INTERACTIONS

As has been suggested, there appears to be a formal standard in epidemiology that if there is something left over after statistical adjustments of racial comparisons, it must be a generalized attribute of race. This is analogous to what philosophers call an "essentialist" concept.[16] There is a generalized view that the force of some fundamental process underlies the exposure-disease relationship and functions as an essential matrix of the individual. Science tries to make this quantifiable and develops a set of rules and logic to evaluate and understand it. In epidemiology, the usual approach is to say that phenotypic variation, or the individual we can observe, is the product of two inputs–genes and the environment–and these are generally considered to be additive quantities. A certain complement of genes occurring in a certain type of environment together produces phenotypic variation:

phenotypic variation=genes+environment.

When the genetic contribution is trying to be understood but cannot be measured directly, that is, on a molecular basis, the equality is rearranged:

phenotypic variation−environment=genes.

The genetic contribution, which is what is trying to be understood, is isolated on one side and the environmental component subtracted from phenotypic variation. This is part of the process used in epidemiology to give some quantitative measure of the contribution of genes in relation to environment.

As a technical process, this is also intrinsically flawed. Neither the tools nor the statistical methods to do this effectively are available. Objections similar to ones raised previously with other analytic strategies also can be raised here:

1) Clearly, not all of the environmental factors can be identified and measured.
2) Many of them can be measured only imprecisely.
3) Perhaps more importantly, these factors are not all additive.

It is not simply a matter of having A+B, but having an interaction in the form of A×B. That is to say, genes in a particular environment have a different outcome. Therein lies the central conceptual limitation of race as a biologic construct for the study of disease causes: the logic of genetic explanations of disease risk (ie, genetic susceptibility) leads us astray since it tempts us to accept essentialist causes. This acceptance may arguably be even less tenable in the context of polygenic disorders, of which prostate cancer may be an example.[17]

Expression of genetic material is context-dependent. Therefore, when trying to determine the pres-

ence or absence of genetic effects, it is important to measure both the environmental and genetic contributions and account for potential interactions. Unfortunately, gene-environment interactions are extremely difficult to measure, and quantifying them is probably well over the horizon of our current technical capabilities. However, there is now a revolution in biology that allows us to make genetic measurements rather than talk about these as philosophical concepts.[17] This is cause for judicious optimism.

CONCLUSION

Serious technical and conceptual obstacles limit the use of race in the study of disease causes; racial comparisons to test genetic susceptibility hypotheses are particularly weak. That said, there is, at least theoretically, the potential to use race or ethnicity to advance our understanding of the cause(s) of racially variable diseases such as prostate cancer. However, much needed work remains, and numerous issues must be resolved in order to move forward. These include a more comprehensive understanding of the factors that mediate the apparent effect of race, valid measures of those factors, and creative strategies that help overcome the technical and interpretive limitations of statistical adjustment.

Finally, the "grand" theories of race-based genetic susceptibility should be renounced and rigorous criteria established to determine when a trait can be ascribed to some genetic origin. At minimum, these criteria should include: 1) evidence that the gene confers susceptibility in each of the groups being compared, 2) definition of the functional mutations or highly specific haplotypes, and 3) differences in the frequencies of the at-risk alleles between the groups. With the ability to make actual observations, that should be put forward as the only useful information: to talk about genetic effects should be to talk about genes. That is the challenge of the molecular revolution. Therefore, let us find the genes, find out how they work, identify the functional relations, and see if the frequency of mutations varies between groups. Genetic explanations, which are actually what the biologic concept of race is all about, should be about genes and not about vague notions that people are different and have different rates of disease, and that since genes cause disease, the differences between the two must be due to genetics.

Literature Cited

1. Centers for Disease Control and Prevention. *Chronic Disease in Minority Populations.* Atlanta, GA: CDC; 1992.
2. Wagenknect LE, Roseman JM, Alexander WJ. Epidemiology of IDDM in black and white children in Jefferson County, Alabama, 1979-1985. *Diabetes.* 1989;38:629-633.
3. Moore ML, Michielutte R, Meis PJ, Ernest JM, Wells HB, Buescher PA. Etiology of low-birthweight birth: a population-based study. *Prev Med.* 1994;23:793-799.
4. Zahm SH, Fraumeni JF. Racial, ethnic, and gender variation in cancer risk: considerations for future epidemiologic research. *Environ Health Perspect.* 1995;103(suppl 8):283-286.
5. Whittmore AS, Wu AH, Kolonel LN, John EM, Gallagher RP, Howe GR. Family history and prostate cancer risk in black, white, and Asian men in the United States and Canada. *Am J Epidemiol.* 1995;141:732-740.
6. Baquet DR, Horm JW, Gibbs T, Greenwald P. Socioeconomic factors and cancer incidence among blacks and whites. *J Natl Cancer Inst.* 1991;83:551-557.
7. Rothman KJ. *Modern Epidemiology.* Boston, MA: Little, Brown and Co; 1986.
8. Liberatos R, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev.* 1988;10:87-121.
9. 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;8:621-628.
10. Klahr S. The kidney in hypertension: villain and victim. *N Engl J Med.* 1989;320:731-733.
11. Law MR, Frost CD, Wald NJ. By how much does dietary sodium restriction lower blood pressure? III: analysis of data from trials of salt reduction. *BMJ.* 1991;302:819-824.
12. Brancati FL, Whittle JC, Whelton PK, Seidler AJ, Klang MJ. The excess incidence of diabetic end-stage renal disease among blacks: a population-based study of potential explanatory factors. *JAMA.* 1992;268:3079-3084.
13. Coughlin SS, Neaton JD, Sengupta A, Kuller LH. Predictors of mortality from idiopathic dilated cardiomyopathy in 356,222 men screened for the Multiple Risk Factor Intervention Trial. *Am J Epidemiol.* 1994;139:166-172.
14. Wild S, McKeigue P. Cross-sectional analysis of mortality by county of birth in England and Wales, 1970-1992. *BMJ.* 1997;314:705-710.
15. Moul JW, Douglas TH, McCarthy WF, McLeod DG. Black race is an adverse prognostic factor for prostate cancer recurrence following radical prostatectomy in an equal access health care setting. *J Urol.* 1996;155:1667-1673.
16 Huncharek M, Muscat J. Genetic characteristics of prostate cancer. *Cancer Epidemiol Biomarkers Prev.* 1995;4:681-687.
17. Schuler GD, Boguski MS, Stewart EA, et al. A map of the human genome. *Science.* 1996;274:540-546.