

Epidemiologic analysis of racial/ethnic disparities: Some fundamental issues and a cautionary example[☆]

Jay S. Kaufman^{*}

*UNC School of Public Health, Department of Epidemiology, 2104C McGavran-Greenberg Hall, Pittsboro Road,
CB#7435, Chapel Hill, NC 27599-7435, USA*

Available online 14 January 2008

Abstract

Racial/ethnic health disparities are a primary focus of epidemiologic research, encompassing both sociological hypotheses about differential treatment as well as biomedical hypotheses about distinctive etiologic processes that might underlie observed disparities. These two main currents in disparities research are often pitted against one another as opposing paradigms. Despite contentious debate about the balance between these hypotheses in the etiology of existing disparities, one consideration that has been largely ignored is that there are important distinctions in the statistical justifications for these two types of inferences. In this article, I review the foundations of causal inference in etiologic epidemiology as applied to studies of racial/ethnic health disparities. I describe normative applications of quantitative techniques for causal inference as they are practiced in research on discrimination in health care and also for research on innate predisposition. I then show why the latter is an injudicious application of this statistical methodology, and illustrate this point with the example of an influential study in the biomedical literature that purported to demonstrate a lesser response to angiotensin-converting enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Health disparities; Statistical inference; Causality; Etiology; Race/ethnicity

Introduction

Racial/ethnic health disparities are a primary focus of epidemiologic research (Kaufman & Cooper, 2001). This research encompasses both sociological hypotheses, for example about differential treatment (Smedley, Stith, & Nelson, 2003), as well as biomedical

hypothesis about distinctive etiologic processes that might underlie observed disparities, such as genetic predispositions (Cooper, 2004). These two main currents in disparities research are often pitted against one another as opposing paradigms, and they have indisputable political implications that affect the way in which research is conducted, interpreted and debated. For example, some political conservatives have decried research emphasis on differential treatment, arguing that it detracts attention from the individual health behaviors and innate predispositions that contribute importantly to observed disparities (Satel, 2001; Satel & Klick, 2005). In contrast, other researchers have argued that the observed social disparities in health are a necessary

[☆] This work was supported in part by a Robert Wood Johnson Foundation Investigator Award in Health Policy Research. The views expressed imply no endorsement by the Robert Wood Johnson Foundation. Additional support was received from the National Center on Minority Health and Health Disparities (P60 MD000244).

^{*} Tel.: +1 9199667435; fax: +1 9199662089.

E-mail address: jay_kaufman@unc.edu

consequence of political and economic policies and ideologies at the societal level (Levins, 2000; Navarro & Shi, 2001), and have asserted that excessive attention to individual factors such as genetics and behaviors is a distraction motivated by ideological rather than scientific justifications (Goodman, 2000).

Despite the voluminous and contentious debate on this question of the balance between individual and social factors in the etiology of racial/ethnic health disparities, one important consideration that has been largely ignored is that the statistical justification for these two types of inferences is fundamentally distinct. Although data are collected, analyzed and interpreted in relation to both of these general hypotheses, the basis for statistical inference and causal attribution differs in significant ways. This difference can be most readily demonstrated by noting that the hypothesis of social discrimination is one in which the causal process under study is in the mind of the physician (or other decision-maker), in response to some acute stimulus, such as the presentation of a patient whose race is observed. The hypothesis of individual predilection or predisposition, on the other hand, places the causal process within the physiology or psychology of the affected individual. In this case, the individual experiences a behavior or physiologic state because of who they are, a condition that is chronic and fundamentally inalterable.

There are numerous ramifications of this conceptual distinction between the two causal hypotheses, but one notable example is that it is possible to design experimental trials to study discrimination directly, whereas it is not possible to design an experimental study that directly assesses innate predilection or predisposition. Scientists can certainly learn about mechanisms at the level of individual physiology and how these contribute to population-level health disparities, but the process is necessarily inductive. For example, an understanding of the role of hemoglobin S in the etiology of malaria helps to explain population-level disparities in the incidence of sickle-cell anemia between people of West African ancestry in contrast to people of Northern European ancestry. This knowledge did not (and could not) arise from statistical inference in population studies of risk factors and disease, however, but rather from physiological studies and inferences from evolutionary theory (Fix, 2003).

In this article, I review the foundation of causal inference in etiologic epidemiology as it is applied to studies of racial/ethnic health disparities. I will describe normative applications of quantitative techniques for causal inference as they are practiced in research on discrimination in health care and also for research on

innate predisposition. I will then show why the latter is an injudicious application of this statistical methodology, and I will illustrate this point with the example of an influential study in the biomedical literature that purported to demonstrate a lesser response to angiotensin-converting enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction (Exner, Dries, Domanski, & Cohn, 2001).

What epidemiologists do?

Causal inference is fundamentally related to experimentation, which is why the randomized controlled trial (RCT) is widely considered to be the “gold standard” for establishment of causality in biomedicine. In an experimental design such as an RCT, the treatment of interest is assigned or withheld without regard to any measured or unmeasured characteristic of the individual. Therefore, the cumulative experiences of the two groups at the end of the study cannot logically result from any systematic differences in covariate distributions (Rubin, 1974). It is the absence of the alternative explanation of an unmeasured common cause that makes a properly conducted RCT such a compelling argument for a causal effect of treatment expressed in counterfactual terms. That is, the treated and untreated groups can each serve as valid substitute populations for the other’s unobserved (counterfactual) experiences, and an effect measure formed by the contrast between the groups is an unbiased estimate of the true unobservable causal contrast between treated and untreated summary values for the target population of interest (Maldonado & Greenland, 2002).

In observational studies, in contrast with RCTs, there may exist any number of measured or unmeasured quantities (Z) that are associated with (but not affected by) a point-exposure of interest (X) and causally precede the outcome (Y). In general, such variables will act to *confound* the observed relation, meaning that the observed association in the data would not converge to the true causal effect as the sample size becomes infinite (Greenland, Robins, & Pearl, 1999). The true causal effect is the one that would be achieved from an experimental manipulation of X . Formally, if actually forcing X to some specific value x_1 would result in a different probability distribution of Y than if we had forced X to some alternate value x_2 , we say that X has a causal effect on Y , and, furthermore, that the magnitude of this causal effect can be described as some contrast of these values, such as a difference or a ratio. Pearl (2000) developed succinct notation that expresses

such an effect without specifying the form of contrast or the levels to which X is fixed:

$$\Pr(Y = y | \text{SET}[X = x]) \quad (1)$$

This notation then allows us to express confounding as the divergence between any such contrast in experimentally manipulated data and the same contrast in passively observed data:

$$\Pr(Y = y | \text{SET}[X = x]) \neq \Pr(Y = y | X = x) \quad (2)$$

If confounding can be attributed entirely to covariate(s) Z , then adjustment, for example via standardization, allows for the unbiased estimation of the true causal effect:

$$\begin{aligned} \Pr(Y = y | \text{SET}[X = x]) \\ = \sum_z [\Pr(Y = y | X = x, Z = z) \Pr(Z = z)] \end{aligned} \quad (3)$$

where summation is over observed values $Z = z$, and Z is unaffected by X (Rosenbaum, 1984).

This is the basic theoretical foundation of statistical adjustment for observational data, and motivates the approach to all etiologic (as opposed to purely descriptive) analyses. The appeal to experimental manipulation for a definition of confounding is essential in order to tie the results back to the real world in some way. That is, the observed data are already a description of the world as it exists factually. Statistical adjustment creates a new set of numbers which do *not* pertain to the factual world, and therefore one must ask: to what world do they pertain? The answer is that they describe distributions of outcome Y under hypothetical manipulations of the exposure, letting all other quantities run their natural course (Greenland, 2005).

The experimental model applied to disparities arising from discrimination

Applying the analytic epidemiologic model described above to health disparities arising from differential treatment is straightforward because the experimental intervention is well-defined. Consider a patient's socially recognized racial identity in relation to some medical diagnosis, evaluation or procedure, such as referral for right-heart catheterization (Schulman et al., 1999). What we want to know is whether the causal process to refer or not refer, which takes place in the physician's mind, is affected (consciously or unconsciously) by the patient's race. In the conduct of an experiment, we can hold all other factors

constant. That is, we can present a decision-maker with two patients who are identical in all relevant respects except for the one characteristic of interest. If the physician's decision is consistently associated with patient's race over a large number of such matched pairs in this experimental design, then the only plausible interpretation is that the decision-maker is using race in order to decide how to act. If there is no good evidence base for considering race in this way, then the decision-making process is discriminatory in the pejorative sense of the word.

A large number of experimental trials with this basic design has been conducted over the years in order to document the role that patient demographics play in clinical decision-making. For example, Loring and Powell (1988) constructed two psychiatric case presentations that were intended to represent undifferentiated schizophrenia, and these were randomly assigned to be from one of five categories, consisting of each of four race/sex combinations (black/white and male/female) and a fifth group lacking any demographic information. These profiles were then assigned a diagnosis by 290 psychiatrists who returned questionnaires through the mail. Even though the case vignettes were identical except for the demographic information, the proportions of various diagnoses varied substantially by randomly assigned sex and race of the hypothetical patients. For example, black patients, especially black men, were much more likely to be assigned the diagnosis of paranoid schizophrenia, indicating that clinicians perceived in these descriptions greater degrees of violence, suspiciousness and dangerousness than for the identical white patients. Furthermore, of 19 respondents who said that insufficient information was available to form a judgment, 12 had vignettes for which race and gender had not been provided.

Many other experimental trials of this basic design have been conducted over the years to demonstrate the causal role of race in decision-making in a variety of settings, usually in situations in which there is general agreement that conditioning the decision on race has no rational justification. For example, Bertrand and Mullainathan (2004) sent fictitious resumes to help-wanted ads in Boston and Chicago newspapers, randomly assigning African-American- or white-sounding names to the otherwise identical applications. In this study, white names received 50% more callbacks for interviews. What these experimental studies have in common is the assurance that there are no variables, measured or unmeasured, that are correlated with the exposure (race) and independently predictive of the outcome. The patients or job applicants can have no

such characteristics because, in fact, they don't exist. They are represented only by a written case presentation or job application that is known to be invariant. Any perceived differences in the outcome distribution, therefore, are attributable to the decision-makers' imposition of racial stereotypes or prejudices onto the imaginary study subjects, which is exactly the effect that one wishes to isolate and quantify.

The analytic epidemiologic model applied to disparities arising from discrimination

In the experimental designs described above, the causal effect of interest is readily defined:

$$\Pr(Y = y | \text{SET}[\text{Race} = r]) \quad (4)$$

and the “setting” action is literal, because the case vignettes or job applications are physically manipulated. With this model in mind, however, it is easy to extend the same logic to observational studies in which no actual manipulation is achieved, but for which statistical manipulation of the observed data is relied upon as a method for estimating what would happen in an experimental scenario. If there are covariates Z that are associated with (but not affected by) race and which are independently predictive of the outcome, these must be measured and adjusted for in order to have confidence that the association measure in the study has a causal interpretation. Specifically, this causal interpretation is the outcome distribution contrast that would be observed under a randomization of race to the case presentations, rather than contrasting individuals based on their observed races.

For example, [Todd, Deaton, D'Adamo, and Goe \(2000\)](#) sought to determine the causal effect of patient race on receipt of analgesics for extremity fractures in hospital emergency rooms. The authors reviewed records from an urban emergency department in Atlanta, selecting for study all black and white patients presenting with new, isolated long-bone fractures over a multi-year study period. They collected all available medical and demographic information on these patients along with their recorded level of analgesic administration, and used regression modeling to adjust simultaneously for these multiple covariates that are potentially associated with race and are predictive of treatment. For example, if black patients were more likely to be female, and if females generally receive more analgesia, then sex would be a potential confounder of the causal effect of interest. In this case, the authors found that after adjustment for all relevant measured covariates, white

patients were significantly more likely than black patients to receive analgesics, despite similar records of pain complaints in the medical record. The risk of receiving no analgesic at all was 66% greater for black patients than for white patients, an effect that could not be explained by any confounding factor known to the investigators.

A less secure basis for causal inference in the observational study of discrimination, in contrast to the experimental designs described above, arises because it is always possible that there is some variable which is known to the decision-maker but not to the analyst, and which creates a spurious association between race and outcome. For example, if black patients in the [Todd et al.](#) study were more likely to have some legitimate counter-indication for analgesia that was not recorded in the medical record, then it would appear to the analyst that emergency room physicians were acting irrationally, when in fact their actions were justifiably motivated by this hidden variable. This is the “fundamental problem of causal inference” that plagues all observational research ([Holland, 1986](#)), and the only general solution is to have sufficient subject-matter knowledge and sufficiently good data that one can have confidence that no confounder of any importance remains unmeasured.

[Heckman \(1998\)](#) notes that even if all relevant covariates are measured, other modeling improprieties, such as incorrect specification of the model form, can lead to bias in causal estimation. Furthermore, if the decision is based in part on an unmeasured characteristic, and if it is obtained by exceeding some critical threshold value, then changes in the distribution of the unmeasured characteristic can also lead to observed inequality in the outcome even in the absence of discrimination. This argument was in fact made recently by former Harvard University President Lawrence Summers in his famously controversial speech about gender inequality in science and engineering ([Fogg, 2005](#)). For example, consider a hypothetical study of gender discrimination in tenure decisions for women scientists at Harvard. Suppose that some unmeasured aspect of mathematical ability has the same mean value in men and in women, but a larger variance in men. If the tenure decision rests on some absolute threshold of ability, then the longer right-tail of the distribution of this unmeasured trait in men will give men a higher probability of exceeding this value, even when there is no discrimination.

The analytic epidemiologic model applied to disparities arising from innate factors

Many investigators wish to evaluate hypotheses of racial disparity that arise from intrinsic factors,

including genetic differences (Fuller, 2003) as well as differences in innate abilities such as intelligence (Rowe, 2005) or athletic prowess (Marks, 2000). The dilemma that arises is that application of the usual epidemiologic analysis rests, as it did in the previous examples, on the direct analogy of the RCT. What we hope to achieve in an analysis of observational data is the effect estimate that we would have been able to observe in the (hypothetical) RCT that we did not conduct. But the obvious problem is that this (hypothetical) RCT is no longer so readily definable for intrinsic factors (Kaufman & Cooper, 1999). The problem is definitional because the no-confounding criterion in Eq. (2) above is expressed in terms of the physical randomization of the study subjects, even if this manipulation is hypothetical. For intrinsic factors, however, the model loses any semblance of interpretability.

This limitation of observational epidemiology has been noted previously (Holland, 1986; Holland, 2001; Kaufman & Cooper, 1999; Zuberi, 2001), but even if one grants such a definition, there is still the practical problem of adjusting for a potentially impossible number of covariates. What finite set of measurable quantities can we adjust for in Eq. (3) to make the treated and untreated groups (e.g., blacks and whites) each serve as valid substitute populations for the other's unobserved (counterfactual) experiences (Maldonado & Greenland, 2002)? Given the pervasive social distinction made between racial/ethnic groups in racially stratified societies, it seems implausible to suggest that within some definable cross-classifications of covariates, blacks and white might be considered to be exchangeable in all respects except for the exposure (Kaufman, Cooper, & McGee, 1997).

If the basis for causal inference is subverted for non-manipulable (i.e., intrinsic) factors, then how can we speak colloquially about health disparities arising from differential distributions of biologic traits? For example, the age-adjusted 2001 incidence of melanoma in black men in the United States was 1.1/100,000, whereas for white men the incidence was 21.5/100,000, or 20-fold higher (U.S. Cancer Statistics Working Group, 2004). Most people would be comfortable attributing this disparity almost entirely to differences in average skin color between these populations, even though this is a trait that is immutable, and for which no randomized intervention is readily conceivable. Furthermore, we might even speak of a counterfactual in which we imagine what would be the incidence in populations of African origin if they were to have light skin, all other factors being equal, and most people would take a number close to 21.5/

100,000 as a reasonable estimate for this counterfactual. Or to take a somewhat less obvious example, consider the $\Delta F508$ mutation in the CFTR gene, a trait that in homozygotes leads to an inability to transport salt in cells of the lungs called cystic fibrosis (Rowe, Miller, & Sorscher, 2005). US whites have a population prevalence of this mutation of about 1/2800 individuals, whereas for US blacks the mutation occurs in only about 1/35,000 individuals (Phillips, Bishop, Woods, & Elias, 1995). This variant appears to be sufficient (although not necessary) for the development of the cystic fibrosis phenotype, and therefore it seems entirely reasonable to attribute much of the observed racial disparity to differential distribution of this trait at the population level.

What the skin cancer and cystic fibrosis examples demonstrate is that we can make inductive arguments about intrinsic traits as causes of disparities so long as we can assume plausibly that environmental factors are distributed more or less equivalently between the two groups (e.g., ultra-violet radiation and skin cancer in the United States) or that environmental factors are essentially irrelevant (e.g., $\Delta F508$ mutation and cystic fibrosis). In this situation, no covariate adjustment (except perhaps age) is necessary, and the causal inference model represented by Eq. (3) can be expressed simply as:

$$\Pr(Y = y | \text{Race} = r) \quad (5)$$

This is because subject-matter knowledge assures that there are no factors that are both strongly correlated with race and highly independently predictive of the outcome, and therefore no “setting” (i.e., physical manipulation) of the exposure is necessary. If we were to consider cystic fibrosis *mortality* as opposed to *incidence*, however, then this *prima facie* no-confounding assumption would no longer be tenable, because socioeconomic and other factors that are associated with race can clearly contribute to disease severity and medical care, which in turn affect mortality (Britton, 1989; O'Connor et al., 2002).

For racial disparities in common multi-factorial disorders such as cardiovascular disease and common cancers, the inductive inferential process exemplified by the cystic fibrosis incidence example does not readily apply. If there is any single sufficient cause for diseases such as diabetes or hypertension, it remains unknown, and most researchers believe that a large number of genetic and environmental factors contribute interactively to risk, each with a relatively small effect. Differential prevalence of some candidate gene would

therefore provide little information about differential prevalence of disease phenotype, since multiple pathways can lead to the same observed disease endpoint. Moreover, we know that contrasting social and economic environments lead to wide variations in phenotype even for groups with common ancestry, such as the comparison of diabetes and hypertension risk across the diverse social environments of the West African Diaspora (Cooper et al., 1997). When aspects of the physical or social environment affect disease risk, and when these are correlated with race, then the effect of race will be confounded, and this confounding cannot be eliminated for both definitional and practical reasons. The definitional reason is that the criteria for no-confounding (Eq. (2)) involves a model of physical manipulation of the exposure that has no interpretability when race is the exposure. The practical reason is that in a racially stratified society, in which race is correlated with virtually all cultural traits, social interactions and economic options, the dimensionality of covariate vector Z in Eq. (3) is impossibly large. That is, one could never hope to measure sufficiently many characteristics of study subjects to make the two groups exchangeable (i.e., valid substitute populations for the others' average counterfactual outcome experiences).

The conclusion regarding the epidemiologic approach to observational data on racial health disparities seems unambiguous. When the investigator's causal hypothesis involves unjustified discrimination on the part of a decision-maker, then the statistical approach to causal inference is generally defensible. The exposure effect can be theoretically randomized, which provides a meaningful standard of what it would mean for the effect to be unconfounded. Furthermore, when the data are observed rather than experimental, so that adjustment for covariates is necessary, the universe of potentially confounding variables is often tractable. This is because the causal process involves the deliberations of a decision-maker, and so the factors that impinge upon this process can themselves be studied and enumerated. On the other hand, when the investigator's causal hypothesis involves an innate characteristic of the study subject, such as genetics, physiology or psychology, then the statistical approach to causal inference on the basis of multivariate adjustment is difficult to justify. There is no meaning to the randomization model in Eq. (1), and therefore the no-confounding criterion in Eq. (2) remains undefined. Furthermore, since there is no hope in the foreseeable future of knowing all of the social and environmental mechanisms that affect risk of chronic diseases such as hypertension and diabetes, and since racism creates imbalances in an

essentially infinite set of social and material factors, no adjustment is plausibly adequate. In common practice, measured adjustment sets are generally so limited that the assertion of causal inference in this context cannot be considered credible.

An example of the analytic epidemiologic model applied to innate factors

In 2001, the *New England Journal of Medicine* published an article by Canadian cardiologist Derek Exner, who had collaborated with Jay Cohn of the University of Minnesota to re-analyze the combined arms of the "Studies of Left Ventricular Dysfunction" (SOLVD) trials (Exner et al., 2001). This secondary analysis compared the efficacy of the angiotensin-converting enzyme (ACE) inhibitor enalapril with placebo in black and white heart failure patients. Exner et al. concluded that current therapeutic recommendations apply to white patients but not necessarily to black patients, and therefore that future therapeutic recommendations should be tailored according to racial background.

This article has proven quite influential, having been cited nearly 200 times in the peer-reviewed literature as of the end of 2006, as documented by *Journal Citation Reports* (JCR), a database published by the Institute for Scientific Information (ISI). Specifically, the article has been cited heavily by proponents of race-specific therapies, as evidence of differential response due to innate differences in physiology of disease between blacks and whites (e.g., Cohn, 2002; Cohn, Loscalzo, & Franciosa, 2003). Moreover, this article has been cited in treatment guidelines as justification for discouraging use of ACE inhibitors in black patients (e.g., Khan et al., 2004). It was also cited by the US Food and Drug Administration (FDA) to justify the new policy that clinical trial data must be reported by participants' race (Haga & Venter, 2003).

Given this substantial impact, it is reasonable to question whether the logic of the analytic exercise conducted in the Exner et al. article is valid in relation to the critique discussed above. The authors stress the randomized design of the SOLVD trials, as though this would confer to them some advantage in obtaining the desired inference. But this appeal to randomization is either misguided or disingenuous, since the focus of the analyses in this article is not on the average treatment effect of enalapril (i.e., the focus is not on the factor that was randomized), but rather on the effect measure modification of this average treatment effect by another variable, race, which was not randomly assigned. Furthermore, there is much prior information available

which suggests that environmental determinants of disease are not well balanced between the racial groups.

The authors began their report by noting that a previously published analysis of the same SOLVD data showed a smaller estimated treatment effect in black than in white patients for two outcomes (death from any cause or hospitalization for heart failure), but that this differential treatment effect did not attain statistical significance after adjustment for measured covariates ($p = 0.08$) (Dries et al., 1999). The authors, therefore, set out to match each of the black subjects with up to four white subjects on several measured factors: enrollment in either the prevention or the treatment trial, ejection fraction, assigned therapy (enalapril or placebo), sex, and age in three broad groupings. The authors claimed that this matching strategy would increase their statistical power for the comparison of treatment effects by race, but this assertion is in fact questionable because race was not randomized (Greenland & Morgenstern, 1990). In cohort studies, matching is primarily indicated in order to reduce costs in the collection of data. Once the data are already collected, however, one can't generally do better by throwing away a large proportion of these data, as the SOLVD analysts did in this paper. This matching strategy could be justified if the authors suspected effect measure heterogeneity, because in this case the additional white subjects would be "off the support" for the causal comparison of interest (Manski, 1993). In this case, however, the results would be generalizable only to the range of values represented by the matching regime (e.g., ejection fractions at which both black and white subjects were observed). The authors made no such restriction in their interpretation, however.

A total of 6797 participants met the inclusion criteria for the study, 4228 from the prevention trial and 2569 from the treatment trial. Of these, 800 categorized themselves as black. Another 5719 participants categorized themselves as white and thus were eligible to be matched with one of the black participants. Of the 5719 white participants, 1196 (21%) were matched with the black patients; the remaining 79% of white patients were discarded. For 579 (72%) of the black participants there was only a single white match, and for 14 black participants (2%), no suitable white match was found.

Patients were randomly assigned to receive up to 20 mg of enalapril or placebo daily, a dosage that could be modified by individual physicians who naturally were not blinded to the patient's race. The analysts collected a number of covariates, including medical care factors such as other drug therapy prescriptions,

and social factors such as the presence of financial distress during the 12 months before enrollment and the highest level of education attained. The primary outcome measures were deaths from any cause and hospitalization for heart failure.

Compared to matched white patients, black patients were in general younger, had higher mean blood pressure, more exposure to recent financial distress, and lower average attained educational level. They were also much less likely to be using aspirin, beta-blockers, or anti-arrhythmics. In light of substantial differences in other dimensions of health and well-being that exist in the wider society, there is no doubt that many other unmeasured variables also differed substantially between the two groups. The matching procedure used by the authors reduced, but did not eliminate, differences in means for observed variables. For example, the relative probability of taking aspirin comparing white to black patients was reduced from 1.88 overall to 1.73 in the matched participants. It is therefore certain that substantial differences remained between the matched participants in a host of other unmeasured variables, not to mention residual confounding due to categorization of the measured variables (Kaufman et al., 1997). For example, 37% of black participants and 24% of matched white participants reported experiencing financial distress before entry into the study, but there is no basis to believe that blacks and white experienced similar levels of deprivation, stress or hardship within the broadly defined category of financial distress.

An important observation is that the absolute risk of the outcomes was higher for black than for matched white participants. That is, even after matching on ejection fraction and other factors, the black group was sicker on average than the white group. For example, 256 (33%) of the matched black participants died, compared to 311 (26%) of the matched white participants (relative risk = 1.25, 95% CI 1.09, 1.44). The discrepancy in baseline risk was even more extreme in the case of the hospitalization, which affected 238 matched black participants (30%) and 226 (19%) matched white participants (relative risk = 1.60, 95% CI 1.37, 1.88). Nonetheless, the authors ignored this discrepancy in baseline risk and focused on the observation that there was an adjusted 49% reduction (95% CI 30%, 63%) in the risk of heart failure hospitalization for the treated group of matched white patients, whereas among the black patients the adjusted risk reduction was only 14%, and not significantly greater than 0 at the $p < 0.05$ criterion.

It is well appreciated in the theoretical epidemiologic literature that groups with higher baseline risk will in general have more modest response to treatment when ratio measures of effect are employed, as they were in this analysis (Maldonado & Greenland, 2002). This is because the counterfactual ratio that defines that causal effect includes in both the numerator and the denominator the proportion of the population that would experience the outcome under either treatment assignment. In this application, for instance, there is some proportion of the participants p_1 who will be hospitalized for heart failure regardless of whether they receive enalapril or placebo, there is a proportion of the participants p_2 who will be hospitalized for heart failure only if they receive placebo, and there is a remaining proportion of the participants p_3 who will not be hospitalized for heart failure regardless of which treatment they receive. A treatment effect is the contrast between the outcome proportion if the entire study cohort were treated with placebo and the outcome proportion if the entire study cohort were treated with enalapril. This contrast may be constructed as the difference measure $[(p_1 + p_2) - (p_1)] = p_2$, which reflects the fact that those participants susceptible to a treatment effect are those in the p_2 group only. A ratio measure of effect, such as that used by these authors, however, takes the contrast as $[(p_1 + p_2)/p_1]$, so that those doomed to be hospitalized under either treatment are no longer cancelled out of the treatment effect measure. For the ratio contrast, an increase in the cohort proportion p_1 necessarily forces the measure closer to the null, even if the susceptible population in the p_2 group is held constant.

Returning to the Exner et al. analysis, it is clear that a population group with higher baseline risk because they are sicker can be characterized as having a larger proportion p_1 . Knowing that the sicker black study population has higher baseline risk of hospitalization than for whites, therefore, it would be possible to predict *a priori* that the ratio measure of treatment effect for this group will be constrained to be closer to the null (White & Elbourne, 2005). Membership in the p_1 sub-population simply represents hospitalization through some causal mechanism that is not affected by enalapril. For example, over half of all heart failure hospitalizations result from excessive sodium retention that precipitates volume overload (Bennett et al., 1998). Patients with good educational and social support resources may successfully avoid hospitalization by regular self-weighing and by then adjusting diuretic dosages in response to fluctuations (Smith, Fabbri, Pai, Ferry, & Heywood, 1997). If the social support and patient–physician communication factors that facilitate this successful self-management are

unmeasured, as they are in the SOLVD data, then they manifest as a reduced treatment effect simply by inflating the proportion p_1 in the cohort sub-population.

Exner et al. do remark in their discussion that “[i]t is also possible that the findings may have resulted from differences between the groups in compliance, diet, medical follow-up, or access to care.” (p. 1356) Nonetheless, they then go on to ignore these caveats and conclude that “the overall population of black patients with heart failure may be underserved by current therapeutic recommendations....[I]t seems appropriate to consider current therapeutic recommendations as applying to white patients but not necessarily to black patients.” The report ends with the suggestion that “therapeutic recommendations may need to be tailored according to racial background.” (p. 1357).

A critique of the Exner et al. article as an example of the analytic epidemiologic model applied to innate factors

The Exner et al. article highlights the inherent problems of attempting statistical inferences about innate factors in observational data. The analysis corresponds to no randomized trial that could be described, even hypothetically, because the causal inference desired is the contrast between a treatment effect for blacks and a treatment effect for whites under the premise that groups are balanced on all important causes of the outcome. This premise motivated the crude matching strategy, but the small number of factors matched and the large number of strong unmeasured causes makes this strategy dubious at best. The authors list a few of the important unmeasured variables that presumably remained unbalanced between the two groups, but the stated conclusions appear untempered by this concern, and the article is widely cited as having demonstrated a universally reduced capacity for therapeutic response among blacks as if the assumption were satisfied. Conditioning on measured variables in the analysis, in addition to matching, is a further attempt to create exchangeability between the race groups. For example, the authors adjusted for the binary variable representing financial hardship, but the adjustment would successfully remove the confounding bias due to this variable only if blacks and whites who reported having experienced financial hardship were equivalent with respect to economic and social factors that might relate to risk of hospitalization. Given the clear violation of this assumption demonstrated in extant demographic data, the equality represented in expression (3) above cannot be considered credible, even as an approximation.

The discussion above noted that in practice we often allow inductive arguments about intrinsic traits as causes of disparities as long as we can assume plausibly that environmental factors are distributed more or less equivalently between the two groups. In the Exner et al. example, however, this is not the case, since psycho-social and economic factors are known to be unbalanced and highly relevant for a sociological outcome such as hospitalization. The larger baseline difference for this outcome variable makes it even clearer that unmeasured factors that differ between blacks and whites in general are important in determining the attained value. Nonetheless, even though there were two primary outcomes defined, and hospitalization was the weaker of the two (in the sense that baseline risk was substantially higher for blacks and assignment was made without the decision-maker being blinded to patient race), the null finding for mortality is largely ignored when the paper is cited as evidence of differential treatment efficacy. Furthermore, the null finding for the harder endpoint of mortality has been replicated elsewhere (Shekelle et al., 2003), and re-analysis of the SOLVD data has shown that hospitalization is unique among possible endpoints in demonstrating a treatment effect differential (Dries, Strong, Cooper, & Drazner, 2002).

In summary, the analytic strategy pursued in the Exner et al. paper provides very little insight into the nature of the observed disparities. The authors' etiologic hypothesis is clearly stated as a physiologic difference between blacks and whites, and yet the analytic epidemiologic model applied to innate factors is dubious for the reasons described above. In this specific application, however, the situation is made even more tenuous by the selection of an outcome for which risk in the unexposed is the most discrepant of all defined outcomes, and for which assignment is made subjectively and without being blinded to patient race. These problems are compounded by an unjustified matching strategy that tosses out about 70% of the available data. The result is a finding that has little inferential value for either supporting or refuting the physiologic hypothesis of interest. Moreover, the observational analysis is designed to mimic a randomized trial that cannot be defined in practical terms. Even allowing that the motivating hypothetical trial remains indefinable, the alternate hypothesis that unmeasured determinants of the outcome are imbalanced between the groups is so likely, and so plausibly consequential, that there is little credibility associated with the assertion that the results of this analysis approximate what would be obtained in the imaginary trial.

The weak basis for inference in the Exner et al. paper must be contrasted not only with its appearance as the lead article in the most prestigious American medical journal, but also with its considerable influence on the field. It is over-interpreted not only as evidence that "black patients with chronic heart failure ...derive less benefit from angiotensin-converting enzyme inhibitors, on average, than do whites" (Satel & Klick, 2005), but also to support the general notion that "...there is now growing evidence for genetic factors being responsible for individual response to therapy." (Hovind, Tarnow, Rossing, Carstensen, & Parving, 2004). Indeed, Bond et al. (2004) cite the article to support the assertion "[R]acial differences in responses to angiotensin converting enzyme inhibitors are thought to result from both genetic and environmental factors." Recall that the Exner et al. article in fact involved no analysis of genetic variants whatsoever, and yet this rather surprising extrapolation is characteristic of many of the citing articles. For example, in an article on racial variation in the prevalence of atrial fibrillation among heart failure patients, Ruo, Capra, Jensvold, and Go (2004) cite Exner et al. to support the statement "Another possible explanation for the lower prevalence of atrial fibrillation in African-Americans than in Caucasians with heart failure may be intrinsic racial differences in atrial membrane stability, atrial conduction pathways, or genetic polymorphisms leading to different susceptibility to the development of atrial fibrillation. For example, polymorphisms have been found to be associated with racial differences in... response to treatment for heart failure." Moreover, the Exner et al. article has also been interpreted as extending this logic to other completely unrelated conditions. Hughes, Moreland, and Bridges (2002), for example, review genetic influences on rheumatoid arthritis in blacks, noting that racial variation in therapeutic response for this condition has never been observed. Nonetheless, they cite the Exner et al. analysis as evidence that such racial variation in treatment response might reasonably be expected.

The application of statistical reasoning in epidemiology has a clear foundation, rooted in the notion of randomization that underlies all models, tests and quantitative inferences (Greenland, 1990). Observational epidemiology makes use of this paradigm by analogy, arguing that we may at times be able to mimic the process of physical randomization that occurs in trials through statistical adjustment. That is, conditional on some measured factors, we can assert that no important unmeasured determinants of the outcome remain substantially unbalanced between treatment groups.

This analogy has often proven quite useful, and has led to insights of enormous public health importance over the previous 50 years, such as the discovery of the causal association between cigarettes smoking and lung cancer. There are settings in which this analogy does not hold, however, and these settings include the attempt to discover innate physiological differences between racial/ethnic groups. The discussion above explains why analytic epidemiology can be used sensibly and successfully to identify discrimination in treatment as a cause of racial disparities in health, but not for the identification of innate predispositions as a cause of racial disparities in health. The article by Exner et al. is an example of how such an analysis can be largely unhelpful for evaluating the hypothesis of interest, and in fact even potentially harmful. Despite its questionable inferential validity, however, this article has proven quite influential, and has been cited to support a large number of claims about racial predisposition to disease, many of which have very little to do with the actual content of the article. This apparent eagerness to embrace the message of racial essentialism therefore seems to represent a very strong prior belief on the part of many researchers. Until this strong predilection for racial essentialism in biological thinking abates, there would seem to be little hope that a more sensible and honest approach to statistical inference in observational data will take hold more widely.

References

- Bennett, S. J., Huster, G. A., Baker, S. L., Milgrom, L. B., Kirchgassner, A., & Birt, J., et al. (1998). Characterization of the precipitants of hospitalization for heart failure decompensation. *American Journal of Critical Care*, 7(3), 168–174.
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013.
- Bond, V., Millis, R. M., Adams, R. G., Williams, D., Obisesan, T. O., & Oke, L. M., et al. (2004). Normal exercise blood pressure response in African-American women with parental history of hypertension. *American Journal of Medical Science*, 328(2), 78–83.
- Britton, J. R. (1989). Effects of social class, sex, and region of residence on age at death from cystic fibrosis. *British Medical Journal*, 298(6672), 483–487.
- Cohn, J. N. (2002). Contemporary treatment of heart failure: is there adequate evidence to support a unique strategy for African-Americans? Pro position. *Current Hypertension Reports*, 4(4), 307–310.
- Cohn, J. N., Loscalzo, J., & Franciosa, J. A. (2003). Nitric oxide's role in heart failure: pathophysiology and treatment — introduction. *Journal of Cardiac Failure*, 9(5), S197–S198.
- Cooper, R. S. (2004). The role of genetics in ethnic disparities in health. In N. B. Anderson, R. A. Bulatao, & B. Cohen (Eds.), *Critical perspectives on racial and ethnic differences in health in late life* (pp. 269–309). Washington, DC: National Academy of Sciences/National Research Council Press.
- Cooper, R. S., Rotimi, C., Ataman, S., McGee, D., Osotimehin, B., & Kadiro, S., et al. (1997). Hypertension prevalence in seven populations of African origin. *American Journal of Public Health*, 87, 160–168.
- Dries, D. L., Exner, D. V., Gersh, B. J., Cooper, H. A., Carson, P. E., & Domanski, M. J. (1999). Racial differences in the outcome of left ventricular dysfunction. *The New England Journal of Medicine*, 340(8), 609–616.
- Dries, D. L., Strong, M. H., Cooper, R. S., & Drazner, M. H. (2002). Efficacy of angiotensin-converting enzyme inhibition in reducing progression from asymptomatic left ventricular dysfunction to symptomatic heart failure in black and white patients. *Journal of American College of Cardiology*, 40(2), 311–317.
- Exner, D. V., Dries, D. L., Domanski, M. J., & Cohn, J. N. (2001). Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. *The New England Journal of Medicine*, 344(18), 1351–1357.
- Fix, A. G. (2003). Simulating hemoglobin history. *Human Biology*, 75(4), 607–618.
- Fogg, P. (2005). Harvard's president wonders aloud about women in science and math. *The Chronicle of Higher Education*, 51(21), A12. <<http://www.president.harvard.edu/speeches/2005/nber.html>> See full text of Summers' remarks at: (Accessed 04.01.07).
- Fuller, K. E. (2003). Health disparities: reframing the problem. *Medical Science Monitor*, 9(3), SR9–SR15.
- Goodman, A. H. (2000). Why genes don't count (for racial differences in health). *American Journal of Public Health*, 90(11), 1699–1702.
- Greenland, S. (1990). Randomization, statistics, and causal inference. *Epidemiology*, 1(6), 421–429.
- Greenland, S. (2005). Epidemiologic measures and policy formulation: lessons from potential outcomes. *Emerging Themes in Epidemiology*, 2, 5. <<http://www.ete-online.com/content/2/1/5>> (Accessed 04.01.07).
- Greenland, S., & Morgenstern, H. (1990). Matching and efficiency in cohort studies. *American Journal of Epidemiology*, 131(1), 151–159.
- Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, 14, 29–46.
- Haga, S. B., & Venter, J. C. (2003). Genetics. FDA races in wrong direction. *Science*, 301(5632), 466.
- Heckman, J. J. (1998). Detecting discrimination. *The Journal of Economic Perspectives*, 12(2), 101–116.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of American Statistical Association*, 81(396), 945–970.
- Holland, P. W. (2001). The false linking of race and causality: lessons from standardized testing. *Race and Society*, 4(2), 219–233.
- Hovind, P., Tarnow, L., Rossing, P., Carstensen, B., & Parving, H. H. (2004). Improved survival in patients obtaining remission of nephrotic range albuminuria in diabetic nephropathy. *Kidney International*, 66(3), 1180–1186.
- Hughes, L. B., Moreland, L. W., & Bridges Jr., S. L. (2002). Genetic influences on rheumatoid arthritis in African Americans. *Immunologic Research*, 26(1–3), 15–26.
- Kaufman, J. S., & Cooper, R. S. (1999). Seeking causal explanations in social epidemiology. *American Journal Epidemiology*, 150(2), 113–120.
- Kaufman, J. S., & Cooper, R. S. (2001). Commentary: considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, 154(4), 291–298.

- Kaufman, J. S., Cooper, R. S., & McGee, D. (1997). Socioeconomic status and health in Blacks and Whites: the problem of residual confounding and the resiliency of race. *Epidemiology*, *6*, 621–628.
- Khan, N. A., McAlister, F. A., Campbell, N. R., Feldman, R. D., Rabkin, S., & Mahon, J., et al. Canadian Hypertension Education Program (2004). The 2004 Canadian recommendations for the management of hypertension: Part II – therapy. *The Canadian Journal of Cardiology*, *20*(1), 41–54.
- Levins, R. (2000). Is capitalism a disease? The crisis in US public health. *Monthly Review*, *51*(4), 8–33.
- Loring, M., & Powell, B. (1988). Gender, race, and DSM-III: a study of the objectivity of psychiatric diagnostic behavior. *Journal of Health and Social Behavior*, *29*(1), 1–22.
- Maldonado, G., & Greenland, S. (2002). Estimating causal effects. *International Journal of Epidemiology*, *31*(2), 422–429.
- Manski, C. F. (1993). Identification problems in the social sciences. *Sociological Methodology*, *23*, 1–56.
- Marks, J. (2000). Review of *Taboo*, by Jon Entine. *Human Biology*, *72*, 1074–1078.
- Navarro, V., & Shi, L. (2001). The political context of social inequalities and health. *Social Science & Medicine*, *52*(3), 481–491.
- O'Connor, G. T., Quinton, H. B., Kahn, R., Robichaud, P., Maddock, J., & Lever, T., et al. Northern New England Cystic Fibrosis Consortium (2002). Case-mix adjustment for evaluation of mortality in cystic fibrosis. *Pediatric Pulmonology*, *33*(2), 99–105.
- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.
- Phillips, O. P., Bishop, C., Woods, D., & Elias, S. (1995). Cystic fibrosis mutations among African Americans in the southeastern United States. *Journal of National Medical Association*, *87*(6), 433–435.
- Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of Royal Statistical Society, Series A*, *147*(5), 656–666.
- Rowe, D. C. (2005). Under the skin: on the impartial treatment of genetic and environmental hypotheses of racial differences. *American Psychologist*, *60*, 163–173.
- Rowe, S. M., Miller, S., & Sorscher, E. J. (2005). Cystic fibrosis. *The New England Journal of Medicine*, *352*(19), 1992–2001.
- Rubin, D. B. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
- Ruo, B., Capra, A. M., Jensvold, N. G., & Go, A. S. (2004). Racial variation in the prevalence of atrial fibrillation among patients with heart failure: the Epidemiology, Practice, Outcomes, and Costs of Heart Failure (EPOCH) study. *Journal of American College of Cardiology*, *43*(3), 429–435.
- Satel, S. (2001). Medicine's race problem. *Policy Review*, *110*, 49–58.
- Satel, S., & Klick, J. (2005). The Institute of Medicine report – too quick to diagnose bias. *Perspectives in Biology and Medicine*, *48*(1), S15–S25.
- Schulman, K. A., Berlin, J. A., Harless, W., Kerner, J. F., Sistrunk, S., & Gersh, B. J., et al. (1999). The effect of race and sex on physicians' recommendations for cardiac catheterization. *The New England Journal of Medicine*, *340*(8), 618–626.
- Shekelle, P. G., Rich, M. W., Morton, S. C., Atkinson, C. S., Tu, W., & Maglione, M., et al. (2003). Efficacy of angiotensin-converting enzyme inhibitors and beta-blockers in the management of left ventricular systolic dysfunction according to race, gender, and diabetic status: a meta-analysis of major clinical trials. *Journal of American College of Cardiology*, *41*(9), 1529–1538.
- Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press.
- Smith, L. E., Fabbri, S. A., Pai, R., Ferry, D., & Heywood, J. T. (1997). Symptomatic improvement and reduced hospitalization for patients attending a cardiomyopathy clinic. *Clinical Cardiology*, *20*(11), 949–954.
- Todd, K. H., Deaton, C., D'Adamo, A. P., & Goe, L. (2000). Ethnicity and analgesic practice. *Annals of Emergency Medicine*, *35*(1), 11–16.
- U.S. Cancer Statistics Working Group. (2004). *United States Cancer Statistics: 1999–2001 incidence and mortality web-based report version*. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute. <<http://www.cdc.gov/cancer/npcr/uscs/>> (Accessed 04.01.07).
- White, I. R., & Elbourne, D. (2005). Assessing subgroup effects with binary data: can the use of different effect measures lead to different conclusions? *BMC Medical Research Methodology*, *5*(1), 15. <<http://www.biomedcentral.com/1471-2288/5/15>> (Accessed 04.01.07).
- Zuberi, T. (2001). *Thicker than blood: An essay on how racial statistics lie*. Minneapolis: University of Minnesota Press.