REASSESSING RACE DISPARITIES IN MARYLAND CAPITAL CASES*

RAYMOND PATERNOSTER
Department of Criminology and Criminal Justice
Maryland Population Research Center
University of Maryland

ROBERT BRAME
Department of Criminal Justice
University of North Carolina–Charlotte

KEYWORDS: capital punishment, racial discrimination, propensity score weighting

A generation of research studies that were conducted in multiple states and covered different time periods has found evidence that individuals who kill white victims encounter a greater risk of facing the death penalty than killers of black victims. More recently, research has also examined the likelihood of death penalty processing for black defendants who kill white victims in comparison with other defendant–victim race groups. In particular, a study in Maryland conducted by Paternoster et al. (2003) found evidence that offenders in black defendant–white victim cases were more likely to be death noticed by prosecutors and to receive a death sentence than other offenders. A recent analysis by Berk, Li, and Hickman (2005) raised questions about some of these findings. In this article, we conduct new analyses and conclude that black defendants who kill white victims face a greater risk of adverse treatment than other types of defendants.

Researchers and policy makers have long been concerned about the prospect of racial discrimination in the imposition of the death penalty in the United States. Historically, racial minorities have been the target of

* We would like to thank Greg Ridgeway, Daniel Nagin, John MacDonald, Richard Berk, and the anonymous reviewers for reading this article and providing invaluable suggestions and comments. Direct correspondence to Raymond Paternoster, Department of Criminology and Criminal Justice, 2220 Lefrak Hall, University of Maryland, College Park, MD 20742 (e-mail: rpaternoster@crim.umd.edu).

unequal treatment at many points in the criminal justice system, from the creation of laws to their enforcement (Ayers, 1984; Berlin, 2003; Davis, 2006; Paternoster, Brame, and Bacon, 2008; Williamson, 1984). One specific manifestation of this unequal treatment in the past has been a disparate impact of the death penalty on African Americans, particularly those who victimized whites. This disparity in how the state imposed the death penalty has been the subject of numerous U.S. Supreme Court challenges, from *Powell v. Alabama* (1932) in 1932 to *Gregg v. Georgia* (428 U.S. 153) and its companion cases in 1976.

The current position of the Supreme Court, announced in *McCleskey v. Kemp* (1987), is that racial discrimination in the imposition of the death penalty is unconstitutional. The Court's majority opinion in *McCleskey* stated, however, that a successful showing of discrimination must demonstrate that unwarranted disparate treatment occurred in a specific case by a specific state actor or actors who possessed an intention to discriminate.[1] Recognizing that the demonstration of intentional discrimination in individual cases is virtually impossible, most research on racial discrimination and the death penalty has examined the prospect of systemic discrimination. That is, researchers try to establish whether cases with similar factual backgrounds and case characteristics receive disparate treatment depending on offender and victim racial characteristics.

Although the U.S. Supreme Court rejected this type of general statistical evidence in *McCleskey*, many individual state jurisdictions and researchers continue to be interested in studying and understanding the issue of systemic disparity.[2] An extensive line of research into racial disparity in the death penalty began after the reinstatement of capital punishment in the states following the *Gregg* decision. In 1990, the U.S. General Accounting Office (1990: 6) published its review of this post-*Gregg* body of research. It concluded,

> In 82 percent of the studies, race of the victim was found to influence the likelihood of being charged with capital murder or receiving the death penalty, i.e., those who murdered whites were found to be more likely to be sentenced to death than those who murdered blacks. This finding was remarkably consistent across data sets, states, data collection methods, and analytic techniques.

1. The Court held that "to prevail under the Equal Protection Clause, McCleskey must prove that the decision makers in his case acted with discriminatory purpose" [*McCleskey v. Kemp* (1987) p. 1766].
2. The Court also concluded that general statistical evidence of the kind presented by McCleskey was best directed at state legislatures rather than at the courts because the former are in a better position to weigh the statistical evidence within the context of local customs and history.

More recently, Baldus and Woodworth (2003) conducted a new review of the capital punishment literature and noted that detailed studies have been carried out in at least 17 jurisdictions (mostly states) since 1990. This more recent research shows that individuals who kill white victims are more likely to be charged with capital murder (death noticed[3]) and death sentenced than those who kill black victims (Baldus and Woodworth, 2003; see also Jacobs and Kent, 2007; Jacobs et al., 2007). This race-of-victim effect seems to be enhanced when black offenders cross racial boundaries and kill a white.

Although the collective weight of the evidence is compelling, an important problem developing in this body of research is that cases in different race groups are unlikely to be entirely comparable with each other. For example, cases that involve black defendants and white victims (BD–WV) will tend to differ on legally relevant case characteristics from cases that involve other racial combinations. If these differences are not adequately adjusted, then comparisons of different race groups on death penalty sentencing outcomes will produce misleading results. Researchers generally address this problem by collecting measures of as many legally relevant case characteristics as possible and then statistically adjusting or controlling for them when comparing different race groups on case outcomes.

Three different approaches to this kind of statistical adjustment have been prominently featured in the general and capital sentencing literatures. The first and most commonly used approach is to estimate a regression model in which disposition is the outcome variable, and race and case characteristics are included as predictor variables. For example, in their study of the prosecutor's charging decision in Maryland, Paternoster et al. (2004) estimated a logistic regression equation that included numerous case characteristics as independent variables along with victim's and offender's race.

A second approach has been described and implemented in the extensive work of David Baldus and his colleagues (Baldus and Woodworth, 2003; Baldus et al., 1998, 2002). This approach involves three steps. The first step is to estimate a regression model that predicts case disposition from a set of legally relevant case characteristics (but the race group variable is not included). The second step is to stratify cases based on their predicted dispositions. The third step is to calculate the differences between dispositions for the different race groups within each stratum and

3. The term "death noticed" is used to describe the process by which a prosecutor informs the defendant that she is seeking the death penalty as a punishment. Under Maryland law, the local prosecutor must file a notification of an intention to seek the death penalty 30 days before trial. The notification must also specify what specific statutory aggravating factors the state will rely on.

then to calculate a weighted average race group difference across the strata.

Yet a third approach has been recently implemented by Richard Berk and his colleagues (Berk and He, 2006; Berk, Li, and Hickman, 2005) and is based on the estimation of two statistical models. The first model used classification and regression trees (CARTs) and random forests to predict race group membership (based on covariates composed of case characteristics). Based on this model, Berk and his colleagues calculated the predicted probability that an individual with a particular set of characteristics is a member of each race group. These predicted probabilities are "estimated propensity scores" (Berk, Li, and Hickman, 2005: 374). The second model then predicts case disposition from the race group variables and the estimated propensity scores. Schonlau (2006) recently adopted a similar approach using federal death penalty data, but his disposition regression models were weighted by a function of the estimated propensity score.[4]

Advantages and disadvantages pertain to each approach. A disadvantage common to all approaches is that they rely on the assumption that no legally important variables have been omitted from the respective estimated model. This critical issue is not, therefore, a statistical problem but a data problem—the researcher must be confident that the major variables related to both race and death penalty decision making have been measured. No statistical "fix" for this "omitted variable bias" problem has been found, and the necessity to compare racial groups with similar case characteristics has been widely recognized since the earliest studies. Nonetheless, policy makers and researchers have continued to collect data and study the issue.

The advantage of the first approach is that regression-based logistic regression models are easily understood by most scholars in the field. At the moment, this claim cannot be made of Berk's random forests and CART approach (Berk and He, 2006; Berk, Li, and Hickman, 2005), which has not yet entered the criminological mainstream and which we cannot say we fully comprehend. Nor can this claim be made for the approach of Baldus and Woodworth (2003), which also has had a limited application within criminology and contains properties that have not been widely discussed nor understood. However, compelling arguments are found in favor of the more general propensity score approach.

4. Unlike Berk and colleagues (Berk and He, 2006; Berk, Li, and Hickman, 2005), Schonlau (2006) does not include the estimated propensity score as a predictor variable in the regression model that predicts disposition. The estimated propensity score is used solely for the purpose of creating weights for a regression model in which race predicts the case disposition. Ridgeway (2006) implements a similar methodology to examine racial disparities in the handling of traffic stops by law enforcement. Our analysis is based largely on Ridgeway's example.

First, once the subgroups of matched cases have been obtained, the diagnosis of covariate balance is transparent. The logistic regression-based approach is not as transparent because it is not clear how effective the statistical controls are for creating comparable racial groups. With the propensity score approach, although the researcher has no basis for concluding that covariate balance is found with respect to unobservables, confidence in overall balance does increase to the extent that important covariates are measured and demonstrably balanced. An additional advantage of the propensity score approach in comparison with regression-based models is that the latter make assumptions about the functional form of the disposition model, whereas the former do not. Finally, propensity score models are also generally well understood in the field, and applications that use the method are now regularly appearing in the published literature (Apel et al., 2007; Ridgeway, 2006; Tita and Ridgeway, 2007).

Perhaps more important, an advantage of the use of propensity score matching procedures is that it puts the researcher squarely on a counterfactual platform, and counterfactuals provide a useful conceptual vehicle for causal thinking about racial disparity (Mackie, 1974; Winship and Morgan, 1999). Racial disparity means that two persons with the same attributes/behaviors are treated differently because of their race (in the case at hand, their race and the race of their victim in combination). As Feinberg and Haviland (2003: 139) have put it, "an action is said to be discriminatory, e.g., with respect to race, if the treatment of an individual would be different had that person been of a different race." Thus, an action is said to be discriminatory if within the capital punishment sentencing system the treatment of a black defendant who kills a white victim is different from the treatment that would be given had the offender's and victim's race been different. The examination of BD–WV racial disparity, then, requires an explicit consideration of the counterfactual—"What would have happened had this case involved some other racial combination of offender and victim?" The counterfactual is, of course, unknown because we only have information about the race of the actual offender and the race of the actual victim. Nevertheless, the conceptual advantage of the propensity score approach is that black offender/white victim cases are compared with similarly situated other cases (at least with respect to observed characteristics). To the extent the estimated propensity scores effectively achieve covariate balance, the two groups of cases are comparable with respect to the probability of "treatment" (a black murdering a white). The propensity score approach, therefore, emphasizes the counterfactual approach to causal thinking and inference.

Although widespread agreement is found about the importance of adjusting for potential differences between race groups when estimating models to identify racial discrimination, some ambiguity is found about exactly how a researcher in this area should proceed. Researchers now have several options to choose from in examining the issue of racial disparity in capital punishment and other criminal justice decisions. It is not clear which option is preferred. This problem was highlighted in a recent article by Berk, Li, and Hickman (2005) that reported the results of a study of death penalty data from the state of Maryland covering the period of 1978–1999. These data had been previously analyzed by Paternoster and his colleagues (2003, 2004). The original study—based on multivariate regression analysis of death-eligible case dispositions—found that black defendants who killed white victims were significantly more likely to be death noticed by prosecutors and death sentenced by juries than other defendants. After their propensity score, CART, and random forests analysis of the Maryland data, Berk, Li, and Hickman (2005: 386), however, concluded that "it is very difficult to find convincing evidence for racial effects in the Maryland data and if there are any, they may not be additive." Moreover, Berk and colleagues questioned the logic of both regression-based models and propensity scores to understand the causal importance of racial discrimination in criminal justice (and we would presume other realms such as employment) decisions. They (2005: 378) take the position that because race is not a variable that can be directly manipulated by the researcher (someone either killed a white or they didn't, and we cannot randomly assign them to kill a white), "[i]t is a stretch to ask what would happened [sic] if a victim's race had been different." They conclude that "[t]he counterfactual needs to be plausible, and here it is not."

In this article, we deal with both the analytical and the conceptual issue of using propensity score models in research on racial disparity. In so doing, we attempt to make both a methodological and a substantive contribution. On the substantive side, we wish to revisit the Paternoster et al. (2003) finding that in Maryland, black offenders who kill white victims are at greater risk of a death sentence. We think the data support this conclusion, but we approach the problem with a somewhat different statistical strategy than traditionally has been applied within this setting. In fact, we think that both the Paternoster et al. (2003) findings and the analysis reported in this article can be reconciled with the findings reported by Berk and colleagues.

On the methodological side, we aim to illustrate the usefulness of the propensity score approach for scholars interested in estimating a great variety of different "treatment effects." We think this approach holds great promise for criminologists, not only for those interested in capital punishment but also for those interested in a wide range of other questions. As part of our effort, we will address Berk et al.'s (Berk and He, 2006; Berk, Li, and Hickman, 2005) position that propensity score methods are causally meaningless when the variable of interest cannot be directly manipulated.

We proceed by first describing the Maryland data used in the Paternoster et al. (2003) and the Berk and colleagues (Berk and He, 2006; Berk, Li, and Hickman, 2005) studies and by then discussing how propensity score methods can address the question of disparate racial impact in the context of capital punishment. Next, we conduct several new analyses with the Maryland data that address the issue of disparate racial impact. We conclude the article with some thoughts about the research and policy implications of our findings.

RACE AS A CAUSAL OR "MANIPULABLE" VARIABLE

It has been argued that it makes no sense to speak about the causal or "treatment" effect of a variable such as race (or gender) that cannot be directly manipulated. Although recently argued by Berk and colleagues (Berk, 1988; Berk, Li, and Hickman, 2005), this view has a long history in the philosophy of science, statistics, and the social sciences (Collingwood, 1940; Freedman, 2000, 2003; Holland, 1986, 2003; Holland and Rubin, 1983; Wright, 1971). The conceptual argument behind this view is that causal questions are at their heart counterfactual questions. The causal effect in the capital punishment process of a black defendant killing a white victim is the difference in outcome between these two cases: 1) the case in which black offender x kills a white victim y and 2) the case that involves other combinations of the race of the defendant and the victim. The inferential problem is that we do not and cannot observe these two different outcomes for the same person; we observe only one outcome for their race and for the race of the person they actually killed. The fundamental counterfactual problem, then, is one of missing data—our inability to observe both outcomes for the same person. In this problem, the "treatment" variable is the combination of defendant's and victim's race.

In a classic experimental study, the treatment condition is randomly assigned. For example, medical researchers randomly place some patients in an experimental group in which they receive a particular drug (the treatment), and they randomly place others in a control group with no treatment. With random assignment of the treatment, the counterfactual is meaningful because persons could have been assigned to either the treatment or the control group. With randomization, individuals treated are on average indistinguishable from those not treated, and a meaningful causal inference can be made about the effect of the treatment. This possibility and meaningfulness of the counterfactual is exactly what is at issue with respect to thinking about race as a causal variable—someone's race cannot be randomly assigned. Because race cannot be directly manipulated, the argument goes, no meaningful counterfactual is found, and we cannot draw any causal inferences about the effect or consequences of race. This strict view is that the only meaningful causal relation is one in which the outcome, y, is produced by the direct manipulation of x, the cause (Wright, 1971).

Our position is somewhat different. We think that although one must be careful about making causal inferences in the absence of direct manipulation, nonexperimental studies can provide valuable information that can inform causal thinking. We do not stand alone in this view. In its discussion of racial discrimination research, the National Academy of Sciences noted (National Research Council, 2004: 79):

> According to a conservative statistical position articulated by Freedman (2003) and others, we cannot draw *any* causal inferences in the absence of manipulability. Thus, viewed as a nonmanipulable attribute, race cannot be said to have a causal effect. Others have suggested that by considering the manipulation of all possible confounders, we can at least create a framework in which causal statements about nonmanipulable variables such as race are possible. This latter position is worth serious explication in the context of measurement of discrimination and is related to ideas set forth in the economics literature going back to Haavelmo in the 1940s.

The position we adopt is the less restrictive one that although causal inference is difficult and certainly hazardous with observational data, the results of observational data can be presented within a framework that permits what the National Academy reports refers to as "causal statements" and what we refer to as causal thinking. Even more important for practical purposes is the issue of conducting "balanced comparisons." One does not have to believe the causal argument to agree that merit is found in comparing outcomes of similarly situated cases in difference race groups. That is, our ultimate goal is to ensure that the comparisons being made are, in fact, fair comparisons (for a related discussion, see Schneider, Zaslavsky, and Epstein, 2004).[5]

At least one of these frameworks is the consideration of thought experiments, for example, about what would happen if a case's race characteristics were to change and about how to measure such a change. Such a thought experiment moves us to consider the value of counterfactuals and propensity score procedures.[6] Although race characteristics cannot be

5. We thank an anonymous reviewer for calling our attention to this article and to this point of emphasis.

6. We would like to point out that although Berk and colleagues (Berk and He,

directly manipulated, we think it is enough that we can think about the possibility of how the observed outcomes might have changed if race characteristics had been different (Wright, 1971). Or, alternatively, we can at least think about how the outcome of one case might compare with the outcome of other similarly situated cases. As the National Academy noted, the usefulness of thought experiments in the social sciences has a long and valued tradition of its own. Over 60 years ago, Haavelmo (1944: 6) argued,

> When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved we have in mind some actual experiment, or some design of an experiment, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes.

Thought experiments, Haavelmo argued, are useful in thinking about causal inference, and the thinking about possible counterfactuals in propensity score matching would be an example of such productive causal thinking.[7]

Moreover, the counterfactual for race is substantively meaningful because although race cannot be experimentally manipulated, the *perception* of race can be and has been randomized. A long history of "audit" or "paired-testing" studies is found in economics (Fix and Struyk, 1993). In these studies, persons of two races (or two genders) with comparable credentials and characteristics are sent to apply for a job or housing, and the difference in their success is taken as an indication of the operation of discrimination. Furthermore, evidence of race (or gender) could be withheld from an evaluator to see what the consequences would be. Beginning in the 1970s and 1980s, U.S. symphony orchestras began to conduct "blind" auditions for positions by having the musician audition behind a screen where their gender could not be detected. Up to that point, female

2006; Berk, Li, and Hickman, 2005) reject the use of causal reasoning with a race "treatment," their propensity score analysis makes an important contribution to the literature in this area—by suggesting the use of a new approach for studying race effects in death penalty sentencing. Greg Ridgeway and his colleagues (Ridgeway, 2006; Ridgeway, McCaffrey, and Morral 2006a, 2006b) at RAND have been developing valuable software for propensity score modeling and have been demonstrating important substantive applications.

7. The point is not that thought experiments should replace "real" experiments; rather, it is whether in the absence of the ability to manipulate the treatment variable anything meaningful can be gained by considering what would happen if that ability were present. Berk (1988: 167) made this point when he conceded that ". . .it is one thing to say that the epistemological status of conceptual experiments may be unclear, and quite another to say that nothing can be learned from them."

musicians were substantially underrepresented (possibly discriminated against) in symphony orchestras. Not all orchestras adopted the screened audition procedures at the same time. Goldin and Rouse (2000) conducted an analysis of the hiring practices of orchestras and found that the use of the "blind" audition significantly increased the probability that a female musician would be hired.

Along similar lines, it is not inconceivable to conduct an actual experiment in which prosecutors would be asked to decide whether they would seek a death sentence after reading a hypothetical case record of the homicide with the perceived race of the defendant and victim experimentally manipulated (for a direct example of this in employment discrimination, see Bertrand and Mullainathan, 2004). In fact, if one were to think seriously about a possible "intervention" here, then a likely candidate would be to keep decision makers in capital cases essentially "blind" with respect to the races of the parties involved much like judges in the aforementioned orchestra auditions.

Under the current federal capital punishment system, local U.S. attorneys make a recommendation to the U.S. Attorney General in Washington, DC, regarding each federal death-eligible case as to whether they think a death sentence should be sought. That recommendation is examined by a review committee along with all supporting legal documents from which all information about the race of the defendant and victim is omitted. The U.S. Attorney General's death penalty review committee is, then, "racially blind." So even if race cannot be directly manipulated, the perception of race certainly can be manipulated, and racial information can be withheld from persons to see what consequences result in the decisions they make so that, in the words of Marini and Singer (1988: 370), we can understand the *responsiveness* of various outcomes to changes in race.

Finally, we would be the first ones to argue that the race of the defendant and of the victim in a capital murder case is not a "cause" of the outcome in the traditional sense. It is, at least, not the same kind of cause as when we think of, say, a prison term causing additional offending compared with probation. Race, like gender, is a social construct that "ha[s] meaning through [its] consequences" (Feinberg and Haviland, 2003: 141). We can, as an empirical task, find out whether different combinations of defendant's and victim's race do have different consequences, and then when convinced that they do, we can address the question as to why that might be. A key element of any such discussion, however, is that the groups being compared are, in fact, similarly situated.

The empirical documentation of a relationship between defendant's and victim's race and the charging or sentencing outcome is but one part of the causal inference. A necessary additional part, which researchers have not, at least in the capital punishment literature, even begun to address, is the telling of a theoretically and historically credible story that explains how this relationship comes about. It is this story that connects the cause with the event, a connective story that the philosopher David Hume has called the "cement of the universe," which is every bit as crucial for causal inference as that which finds a relationship between x and y (Hume, 1938 [1740]; Mackie, 1974).

Our view of causal inference, then, is less conservative and less restrictive than the view that asserts that we cannot draw any causal inferences whatsoever if we cannot manipulate the treatment variable. We think a causal statement/inference can be made in the absence of direct experimental manipulation when reasonable grounds are found for believing that "strong ignorability" in the counterfactual exists (Rosenbaum and Rubin, 1983, 1984a, 1984b). Ultimately, then, the value of observational research with propensity scores or studying what Little and Rubin (2000) and Pearl (2000, 2003) have called *potential outcomes* is based on the assumption that the outcomes of similarly situated cases are being compared. An empirical demonstration of balance (or comparability) with respect to important covariates (between the treatment and control groups) is an important step in this direction. Ultimately, we agree with the view of causality expressed by Marini and Singer (1988) who asserted that the evidence needed to support a causal inference must be considered in light of the subject matter. When experimentation and direct manipulation of the independent variable is not possible in a field, then a greater need is found for replication of studies in different contexts with different methodologies and analytic strategies and for the ability to explain possible causal connections that makes a case of causal inference more or less compelling (Hume's "cement").[8]

Make no mistake; we do not think the more conservative definition is wrong. In this regard, we are actually in agreement with Berk (1988: 159) that "Definitions are neither right nor wrong. The question is whether a definition is useful." We think the definition of causality discussed here is useful. But, even if some scholars do not accept the value of strict causal reasoning in this setting, we still think it is useful to compare the outcomes of similarly situated BD–WV cases and cases with other race characteristics.

8. A causal inference would require these elements: 1) a relationship or covariation between x and y, 2) a temporal order in which x precedes y in time, 3) an absence of spuriousness so that alternative causes of y have been considered, 4) a credible story connecting x with y, and 5) the replication of 1–4 in different contexts with different methodologies.

THE MARYLAND DEATH PENALTY DATA

The Maryland death penalty data originally collected by Paternoster and his colleagues has been described in detail in two earlier works (Paternoster et al., 2003, 2004). Data collection covered offenses that occurred between July 1, 1978 and December 31, 1999. The data collection effort began with the identification of "death eligible" cases from the universe of approximately 6,000 first- and second- degree murders committed in the state of Maryland. Cases were considered to be "death eligible" if *either* 1) the state's attorney filed a notice of intent to seek the death penalty, *or* 2) the facts of the case clearly establish that a first-degree murder was committed, the defendant was the principal in the first degree (or met the state's second-degree exception), the defendant was eligible by age (18 years old or older) at the time of the offense, the defendant had not been diagnosed with mental retardation at the time of the offense, and the murder included at least one statutory aggravating circumstance. Approximately 300 homicide cases were found in which the state's attorney did not file notice of intent to seek the death penalty, and it was not clear to the research team whether the case could be characterized as "death eligible." Each of these cases was submitted to panels of five to ten attorneys (both former prosecutors and members of the defense bar) who had experience in death penalty cases. Based on the ratings of these attorneys,[9] fewer than 50 of the 300 questionable cases were characterized as "death eligible." Overall, the research team could confidently characterize 1,311 cases as "death eligible."

An extensive list of case, offender, and victim characteristics was assembled on each of these 1,311 cases. Unfortunately, information on some of these background characteristics was unavailable for some cases. It is desirable to include as many cases as possible with as many characteristics as possible. But it is not possible to optimize both of these quantities simultaneously. An increase in the number of cases necessitates a loss of some background characteristics, and an increase in the number of background characteristics necessitates a loss of some cases. After examining several different possibilities, a list of 93 variables with complete data on

9. Panel members read detailed narratives of each homicide in question that outlined the facts of the case and all known information about the offense, the offender, and the victim. After reading these narratives, panel attorneys were asked to make these two judgments: 1) "Do you think this case is 'death eligible' under Maryland law?" and 2) on a scale from 1 ("not very confident at all") to 10 ("very confident"), how confident do you feel in making this determination? A case was ultimately considered "death eligible" if most of the panel attorneys said that the case was death eligible and the confidence of that rating averaged 5.0 or higher.

1,202 (91.7 percent) of the original 1,311 cases was assembled.[10] Additionally, 72 cases had missing information on either the defendant's or the victim's race. These deletions left 1,130 (86.2 percent) of the original 1,311 cases available for analysis.[11] Because each of these cases represents a single prosecution, we refer to this database as the "prosecution sample." Some cases within the prosecution sample represent subsequent prosecutions of the same individual offender.[12] To check on the sensitivity of our results, we also analyze a "defendant sample" of 1,041 cases (1,130 prosecutions minus 89 subsequent prosecutions of individuals who had already been prosecuted once) [for discussion of a similar issue in the Georgia study that formed the basis for the *McCleskey v. Kemp* (1987) litigation, see Baldus, Woodworth, and Pulaski, 1990: 44].

Table 1 presents an assessment of race characteristics and case dispositions in the Maryland data. This analysis indicates that black defendants accused of killing white victims (BD–WV cases) are more likely to be death noticed by prosecutors and actually receive a death sentence than other defendant/victim race combinations. Both of these differences are statistically significant at any conventional significance level. As noted, however, BD–WV cases differ in important ways from other types of cases. The result is that the analysis in table 1 is likely to reflect both the influence of defendant/victim race characteristics and the other legitimate factors that are associated with both race and case outcomes. In the next section, we assess the extent to which background characteristics for BD–WV cases differ from those of other cases.

10. We think this list is a comprehensive collection of the most important variables that are likely to explain both the "treatment" (defendant's and victim's race) and the prosecutor's charging decision. We do not think we have omitted any variable that has been found in previous research in this area to be a major consideration (see Baldus, Woodworth, and Pulaski, 1990; Baldus et al., 1998, 2002).

11. The data analyzed in this article are the same as those described in Paternoster et al. (2004) but differ slightly from the data discussed in Paternoster et al. (2003). Specifically, the total number of cases (1,311) and the number of cases that remain after deletion of observations with missing data on non-race variables (1,202) are the same as reported in Paternoster et al. (2003). But analyses of victim and defendant race in Paternoster et al. (2003) were based on a sample of 1,061 cases. After publication of the 2003 report, the research team identified information on victim and defendant race for an additional 69 cases, which increased the sample size to 1,130 cases. As discussed in Paternoster et al. (2004), the substantive conclusions remained the same when these new cases were added to the sample.

12. Successfully appealed cases in which the conviction was vacated, for example, would be sent back to the original jurisdiction where the local state's attorney could decide either to file a new notice to seek death in the retrial or to decline to seek death again.

Table 1. Defendant/Victim Race and Case Dispositions in the Maryland Data (Unweighted Analysis)

Defendant/Victim Race	$n =$	Percent Death Noticed	Percent Death Sentenced
Prosecution Sample	1,130	28.76	6.28
Black D/white V	244	46.31	13.93
Other cases	886	23.93	4.18
Unweighted difference*		22.38	9.75
Defendant Sample	1,041	25.74	4.61
Black D/white V	214	41.12	10.28
Other cases	827	21.77	3.14
Unweighted difference*		19.35	7.14

*$p < .05$

BACKGROUND CHARACTERISTICS

As noted, our database is composed of 93 background characteristics for each of the 1,130 prosecutions and 1,041 unique defendants. A key concern that motivates this analysis is that the BD–WV cases may differ from non-BD–WV cases in ways that call the comparability of the two groups into question. To check on this possibility, we examined the means on each of the 93 covariates for BD–WV cases and other cases in the prosecution sample (appendix A) and the defendant sample (appendix B). One covariate is the county in which the case is prosecuted. This variable is a categorical variable with seven values. Thus, we actually need to compare 99 sets of means between the groups. The first two columns of each appendix display the covariate means for the BD–WV group and the other cases (unweighted), respectively. With the exception of one variable (pvfelony), all covariates are coded 1 if the characteristic is present in the case files and 0 if the characteristic is absent. The number of prior violent felony convictions is coded 0, 1, 2, and 3 or more convictions. For now, we confine our attention to the BD–WV covariate means and to the unweighted non-BD–WV covariate means. These data provide an assessment of between-group comparability in the raw data.

When the means for the two groups are approximately the same on a particular characteristic, we say the groups are "balanced" with respect to that characteristic. Following Rosenbaum and Rubin (1985), we assess balance by calculating the "standardized difference" statistic as implemented by Ridgeway, McCaffrey, and Morral (2006a: 3). In general, for any given covariate, this statistic is calculated by dividing the difference between the means for the treatment and control group by the standard deviation for the treatment group. Rosenbaum and Rubin (1985) suggest that a standardized difference statistic in excess of .20 or less than –.20 indicates lack of balance between the groups on that characteristic.

The standardized differences between the BD–WV cases and the unweighted non-BD–WV cases are listed in the SD(U) column of appendices A and B. Applying the Rosenbaum–Rubin standard to the prosecution sample, we identify 13 imbalances out of the 99 comparisons (dark shading). If we adopt a less stringent standard of using .15 and –.15 as cutoff points, then we identify 24 imbalances out of the 99 comparisons (darker and lighter shading). For the defendant sample, we identify 12 imbalances out of the 99 comparisons using a .20/–.20 cutoff, whereas we find 20 imbalances if we use a .15/–.15 cutoff. Clearly, cases that involve black defendants accused of killing white victims differ in several important ways from other cases.

METHODOLOGY FOR ADJUSTING BETWEEN-GROUP DIFFERENCES

To address this lack of comparability, we use an R program (TWANG—Toolkit for Analysis of Nonequivalent Groups) written by Ridgeway, McCaffrey, and Morral (2006a, 2006b) to accomplish these two objectives: 1) Estimate a statistical model of the probability that a case with a particular set of background characteristics is a BD–WV case (i.e., a propensity score), and 2) create weights based on the propensity score and apply those weights to non-BD–WV cases. The goal of the weighting procedure is to ensure both that non-BD–WV cases with high propensity scores receive more weight and that non-BD–WV cases with low propensity scores receive less weight in the analysis. To investigate whether the propensity score estimates produce a weighted comparison group that is similar to the BD–WV group, the program calculates weighted means for each of the covariates over the subsample of non-BD–WV cases. Then, a weighted standardized difference statistic is calculated to assess whether the propensity score weights have made the non-BD–WV cases more comparable with the BD–WV cases (which remain unweighted). Assuming the weighted comparison group is reasonably comparable with the BD–WV group, we then can estimate the weighted disposition difference between the two groups.

This analysis corresponds to estimators in the literature on treatment effects that produce estimates of the "average effect of treatment on the treated" (Ridgeway, McCaffrey, and Morral, 2006a, 2006b). The accompanying thought experiment involves the following steps: 1) We see what the outcomes of the BD–WV cases are, 2) we want to know what the outcomes of the BD–WV cases would be if they had not been BD–WV cases (the counterfactual), 3) the weighted comparison group outcome mean provides us with an estimate of the counterfactual, and 4) the difference between steps 1 and 3 provides us with an estimate of the difference between what the BD–WV cases actually experienced and what they would have experienced if they had not been BD–WV cases.

The propensity score estimator is based on generalized boosted regression models as discussed in McCaffrey, Ridgeway, and Morral (2004). These models are estimated by minimizing the difference between the covariate means for the BD–WV cases and the comparison cases. The result is a statistical model that predicts the probability that each case is a BD–WV case and that produces weights (applied to the comparison cases) that minimize differences between the BD–WV cases and the comparison cases on background characteristics.[13] With the estimated propensity scores in hand, each BD–WV case receives a weight of 1.0, whereas each comparison case receives a weight equal to the odds of the estimated propensity score (i.e., the estimated propensity score divided by the complement of the estimated propensity score).

COMPARING BD–WV CASES TO WEIGHTED NON-BD–WV CASES ON BACKGROUND CHARACTERISTICS

We estimated the generalized boosted regression model to obtain the estimates of the propensity scores in the prosecution and defendant samples. The program uses the estimated propensity scores to calculate weights that are then used to estimate weighted comparison group means. The weighted comparison group mean then is plugged into the formula for the standardized difference statistic. This statistic depends on three quantities, as follows: 1) the BD–WV mean, 2) the non-BD–WV mean, and 3) the standard deviation of the covariate for the BD–WV group. The first

13. The distribution of the weights from the model is available by request. Essentially, the distribution is skewed, as expected, with most cases that are not black defendant/white victim cases having small weights. This distribution must be true because the effective sample size for the weighted non-BD–WV sample is much smaller than the unweighted sample size. One reviewer expressed concern that large weights might "drive the effect estimation." We did not share this concern in the problem at hand. First, the analyst would not want to omit the cases with large weights because those are the cases that are most like the treatment cases. Second, the price of extreme weights is that they decrease the effective sample size of the comparison group, which increases the standard error for any weighted means from the nontreated group. The cost, then, is less statistical power in a hypothesis test, but this would not seem to be a problem in this study with our fairly large sample size. We thank Greg Ridgeway for his assistance on this point.

and third quantities remain the same as in the unweighted analyses discussed above. For the weighted analyses, however, the second quantity is based on the weighted non-BD–WV mean rather than on the unweighted mean. Thus, the only difference between the weighted and unweighted standardized differences is the use of the weighted and unweighted non-BD–WV covariate means.

Alongside the unweighted comparisons, appendices A and B also present the weighted comparisons for each covariate. Based on the results in both appendices, it seems that the propensity score weighting procedure implemented by the R program was successful in eliminating virtually all imbalances. Appendix A (prosecution sample) shows that one imbalance (victim put in water either before or after killing) remained after weighting, and appendix B (defendant sample) also reveals a single imbalance (defendant implicated in other killings) after weighting. Concern about these imbalances is mitigated by the fact that both are based on characteristics that occur with great rarity in the data (less than 5 percent of the cases have either of these characteristics) for both BD–WV cases and comparison cases.[14]

ANALYSIS RESULTS

We now turn to the question of whether black defendants who kill white victims experience adverse treatment when compared with similarly situated cases with other race characteristics. The unadjusted analyses reported in table 1 clearly show that black defendants who kill white victims are statistically more likely to be death noticed and death sentenced when compared with other cases. The analyses presented in this section, however, involve a comparison between the outcomes of BD–WV cases and the weighted non-BD–WV cases discussed previously. Our analyses are presented for both the prosecution and defendant samples. Table 2 presents the relevant comparisons.

The top panel of table 2 is based on the prosecution sample that compares the 244 BD–WV cases to an effective sample size (ESS) of 361.89 "other cases." The effective sample size is a measure of the number of "other cases" that are reasonably comparable with the BD–WV cases (Ridgeway, 2006: 15). Recall from the top panel of table 1 that 886 comparison cases are found in the prosecution sample. The effective sample

14. We also estimated auxiliary weighted regressions using these marginally imbalanced covariates as predictor variables along with the variable that indicates whether the case is a BD–WV case. These auxiliary regressions yielded estimates of BD–WV disparities of nearly identical magnitude to those reported in table 2.

size statistic of 361.89 implies that many non-BD–WV cases were relatively incomparable with the BD–WV cases. These incomparable cases can be viewed as the source of the many unweighted imbalances reported in appendices A and B. After weighting, the influence of these cases is reduced, but the price for that benefit is reduced effective sample size and reduced statistical power for the between-group comparisons.

Table 2. Defendant/Victim Race and Case Dispositions in the Maryland Data (Weighted Analysis with County Adjustments)

Defendant/Victim Race	$n =$	Percent Death Noticed	Percent Death Sentenced
Prosecution Sample			
Black D/white V	244	46.31	13.93
Other cases	361.89	30.52	4.23
Weighted difference		15.79	9.70
t ratio for weighted difference*		3.72	4.01
Defendant Sample			
Black D/white V	214	41.12	10.28
Other cases	308.14	29.58	3.73
Weighted difference		11.54	6.55
t ratio for weighted difference*		2.51	2.80

NOTES: Sample size for "Other cases" is given by the "Effective sample size" described in the text.
*$p < .05$

Considering the prosecution sample, the unweighted difference in death notice rates between BD–WV cases and other cases was 22.38 percentage points (table 1). After weighting, this difference was significantly reduced to 15.79 percentage points (a 29.4 percent reduction). *But the analysis shows that BD–WV cases are still more likely than other cases to be death noticed after adjusting for between-group differences. The death sentence analysis indicates that BD–WV cases are significantly more likely than other cases to be death sentenced.*[15] Moreover, the weighted analysis results are virtually identical to the unweighted results presented in table 1.

15. The test for statistical significance in each case is a t ratio, which is calculated by dividing the difference between the death notice and death sentence rates for the two groups by the estimated standard error of the difference. A t ratio or t statistic whose absolute value exceeds 1.964 is considered to be statistically significant at the two-tailed $p < .05$ significance level.

We now turn to the defendant sample. In the unweighted analyses, the BD–WV death notice rate was 41.12 percent, whereas for other cases, the death notice rate was 21.77 percent. The difference between these rates is 19.35 percentage points (table 1). After weighting, however, this difference is nearly cut in half. The death notice rate for the weighted comparison cases is 29.58 percent. When we compare this rate with the 41.12 percent rate for BD–WV cases, the difference is 11.54 percentage points (a 39 percent reduction). *Nevertheless, the analysis still suggests that BD–WV cases are treated adversely in comparison with other cases at the death notice decision point.*

The unweighted difference between the death sentence rate for the BD–WV cases (10.28 percent) and the other cases (3.14 percent) is 7.14 percentage points. After weighting, the death sentence rate for non-BD–WV cases is 3.73 percent, which produces a weighted difference of 6.55 percentage points. Thus, the weighted estimate is about 8.3 percent less than the unweighted estimate. *For the defendant sample, then, we conclude that BD–WV cases are more likely to be death sentenced than non-BD–WV cases.*

THE INFLUENCE OF COUNTY ADJUSTMENTS

All analyses reported so far with the Maryland data presume the importance of adjusting for the jurisdiction in which the prosecution was brought. Before weighting, for example, the analysis reveals that the prevalence of Baltimore County cases is significantly higher for BD–WV cases than for other cases, whereas for Baltimore City, the reverse is true. We think it is important to also consider BD–WV disparities in the absence of adjustment for jurisdictions. Consider the following hypothetical example. We have two cases that differ in only these three respects: 1) One case is a BD–WV case, and the other is not, 2) the cases are brought in different jurisdictions, and 3) the cases are processed differently in these jurisdictions (i.e., in one jurisdiction the defendant is death sentenced, whereas in the other jurisdiction the defendant is not death sentenced). In this example, an adjustment for the jurisdiction implies that these two cases are not comparable with each other—although the only material difference between them is the jurisdiction in which the case is brought. But if the cases are comparable with each other, then should we disallow a comparison of them because they are in different jurisdictions? Put differently, should we feel better about race disparities that might exist simply because some of those race disparities result from different prosecutorial practices in different jurisdictions? After considering this issue for the past several years, we have come to the conclusion that jurisdiction differences should not be adjusted in a model that estimates racial disparities for the state of Maryland. To assess the impact of this adjustment on our results, we now present an analysis of the Maryland data without adjusting for the jurisdiction in which the case was brought.[16]

A comparison of BD–WV cases with cases that involved other racial combinations on all characteristics (with the exception of jurisdiction) for both the prosecution and defendant samples revealed substantial covariate imbalance before covariate adjustment with propensity scores. The unweighted comparisons reported are identical to those reported in appendices A and B. The weighted comparisons, however, are not, in general, the same, because the jurisdiction in which the case was brought was an influential contributor to the propensity score weights in the jurisdiction-adjusted analyses reported above.

As in appendices A and B, however, the comparisons indicate substantial improvement in covariate balance after weighting (the results of these comparisons are not reported here but are available from the first author by request). Within the prosecution sample, no covariates had a standardized difference in excess of .20 (or less than –.20) after weighting. One covariate (defendant forced way into place of homicide) had a standardized difference in excess of .15. All other imbalances seem to have been substantially reduced by the weighting procedure. Within the defendant sample, none of the covariates exhibited a standardized difference in excess of .20. Two covariates did have standardized differences in excess of .15 (defendant implicated in other killings and defendant forced way into place of homicide), but the other differences have been substantially reduced because of weighting the non-BD–WV cases.[17] As was true for the analysis in which an adjustment was made by jurisdiction, these results provide confidence that BD–WV cases are being compared with a comparable group of cases (observed covariates are balanced) that involve other racial combinations.

Table 3 presents weighted estimates of disparate treatment of BD–WV cases in Maryland without adjusting for the jurisdiction in which the case was brought. For the prosecution sample, the analysis indicates (as before) that 46.31 percent of BD–WV cases are death noticed by prosecutors.

16. Another, different, question is whether evidence of racial disparity is found within Maryland counties. Unfortunately, in all but a couple of counties because of small numbers, we run into the problem of not enough cases and too many potential independent variables. In a logistic-based regression analysis of Baltimore City and Baltimore County cases separately, black offenders who killed white victims were more likely to be death noticed and death sentenced than offenders in cases that involved other racial combinations.

17. We also estimated auxiliary weighted regressions using these marginally imbalanced covariates as predictor variables along with the variable that indicates whether the case is a BD–WV case. As in our previous analyses, these auxiliary regressions yielded estimates of BD–WV disparities of nearly identical magnitude to those reported in table 3.

whereas 13.93 percent of BD–WV cases are death sentenced. By comparison, 25.79 percent of the weighted non-BD–WV cases are death noticed, and 3.05 percent of these cases are death sentenced. *These estimates imply that BD–WV cases receive adverse decisions at higher rates than the weighted non-BD–WV comparison cases within the prosecution samples.*

The bottom panel of table 3 presents the analysis results for the defendant sample. As before, the analysis reveals that 41.12 percent of BD–WV cases are death noticed by prosecutors, whereas 10.28 percent of BD–WV cases are death sentenced. Within the weighted comparison group, the death notice and death sentence rates are substantially lower. The death notice rate for the weighted non-BD–WV cases is 24.39 percent, whereas the death sentence rate for these cases is 2.93 percent. *For both death notices and death sentences, then, the analysis suggests that black defendants who kill white victims are more likely to receive adverse decisions.*

Table 3. Defendant/Victim Race and Case Dispositions in the Maryland Data (Weighted Analysis without County Adjustment)

Defendant/Victim Race	$n =$	Percent Death Noticed	Percent Death Sentenced
Prosecution Sample			
Black D/white V	244	46.31	13.93
Other cases	424.15	25.79	3.05
Weighted difference		20.52	10.88
t ratio for weighted difference*		5.20	4.71
Defendant Sample			
Black D/white V	214	41.12	10.28
Other cases	395.78	24.39	2.93
Weighted difference		16.73	7.35
t ratio for weighted difference*		4.12	3.28

NOTES: Sample size for "Other cases" is given by the "Effective sample size" described in the text.
*$p < .05$

DISCUSSION AND CONCLUSIONS

Our analysis of the Maryland death penalty data leads us to the conclusion that black defendants who kill white victims (BD–WV cases) are more likely to receive adverse treatment than similarly situated non-BD–WV defendants. To reach this conclusion, we studied unadjusted death notice and death sentence rates for prosecutions and for unique defendants. These analyses suggested strong disparities against BD–WV cases. But because BD–WV cases differ in important ways from non-BD–WV cases, it is important to take these differences into account.

To address this issue, we conducted four sets of analyses adjusted for important differences between the BD–WV cases and the non-BD–WV cases. For each of these four analyses, a set of case weights was constructed to ensure that the weighted non-BD–WV cases were comparable with the BD–WV cases on over 90 case characteristics. Analysis 1 examined death notice and death sentence rates for BD–WV and non-BD–WV prosecutions while adjusting for the jurisdiction in which the case was brought. Analysis 2 was based on a similar comparison with the exception that unique defendants were studied rather than prosecutions (i.e., the same defendant can be prosecuted more than one time). Both of these analyses provide strong evidence that BD–WV cases are more likely to be death noticed and death sentenced than comparable non-BD–WV cases.

Analyses 3 and 4 paralleled Analyses 1 and 2 with the exception that the case weights were not permitted to reflect the influence of the jurisdiction in which the case was brought. We think this issue is important because it may be necessary for policy makers to consider the effects of racial disparities separately from jurisdiction-specific charging and sentencing practices. For example, an analysis that adjusts for the influence of jurisdiction would tend to discount racial disparities that appear because of the charging and sentencing practices of different jurisdictions. We do not argue that this evidence is the only evidence that should be considered in a discussion of racial disparities, but we believe it is relevant. In any event, both of the analyses that avoid weighting for jurisdiction provide strong evidence that death notice and sentencing rates are higher for BD–WV cases than for non-BD–WV cases.

We think these results add to the collective body of research on racial disparities in the nation's capital punishment system. Our results are based on propensity score weighting procedures that adjust for possible alternative explanations differently than regression-based models. We think an important advantage of propensity scores is that the issue of covariate balance is transparent. To the extent that important covariates are balanced, greater confidence can be placed in any conclusion about the effect of a hypothesized "treatment" variable. Another advantage of the propensity score approach is that it is becoming as easily understood and comprehended as standard regression-based approaches. Now numerous applications of propensity score procedures are found in the criminological literature, and the method will soon be in the toolkit of most researchers. Finally, use of propensity scores squarely puts on the table the question of the counterfactual—"What would have been the outcome in this case had the race of the defendant and victim been different?"—which we have argued permits a productive discussion about what consequences race may have for individuals in the capital punishment system.

The substantive finding in our article with propensity score procedures confirms our previously reported results in Paternoster et al. (2003) that black offenders who kill white victims in Maryland over the time period we studied were more likely than offenders in cases that involved other racial combinations to have the local prosecutor charge them with a capital offense given that the case was death eligible. Berk, Li, and Hickman (2005) in their own reanalysis of the Maryland data also reported that cases that involved black offenders and white victims were more likely to be charged with a capital crime than cases that involved other racial combinations (Berk, Li, and Hickman, 2005: Table V, p. 375). The original findings and those reported here differ from Berk, Li, and Hickman's propensity analysis only in their magnitude and statistical significance.[18] In our current and previous work with the Maryland data, we compared black offender and white victim homicides with all others, whereas Berk, Li, and Hickman treated defendant/victim racial combinations as a four-category nominal variable (black defendant/white victim; black defendant/black victim; white defendant/white victim; and white defendant/black victim). Advantages and disadvantages can be found in leaving all four race groups in separate categories. On the one hand, comparisons between black defendant/white victim cases and each of the other groups provides a separate test of the hypothesis that BD–WV cases receive adverse treatment. On the other hand, conducting each of these tests separately means that each test is less powerful than a comparison of BD–WV cases with all other cases combined. We believe that substantive merit is found in comparing BD–WV cases with all others combined because such a comparison represents a more powerful inquiry into the hypothesis that because of the racial threat they possibly present, such cases are treated differently than others. Previous theory and research would support such a prediction a priori (Blalock, 1967; Jacobs and Carmichael, 2002; Kent and Jacobs, 2005; Stults and Baumer, 2007), and we think this is the comparison of greatest interest to policy makers.

We have repeatedly mentioned that the use of propensity score adjustment does not remedy the problem of omitted variable bias that plagues all observational studies, and we note that researchers must take great pains to measure as many important covariates as possible. We do, however, think that with careful and comprehensive measurement and sound

18. Our results remain contrary to Berk, Li, and Hickman's classification and regression tree analysis, which found that a capital charge was more likely in cases that involved a white defendant and a white victim.

statistical analysis, it is possible to employ useful statements about the links between race and case dispositions, although race is not capable of being directly manipulated. Because our position is that in areas where experimental manipulation of the independent variable is not feasible, such as the issue studied in this article, multiple studies are important in building a sound empirical basis for making causal inferences, we think the results we have reported are important. We also, however, think that much additional work needs to be done.

One particularly important area for additional work is the building of a credible story linking *x* with *y*—linking the race of the victim murdered in a homicide with the prosecutor's charging decision. Little critical thinking has been offered about why prosecutors would be more likely to charge capital murder in white victim cases than in black victim cases, and no empirical research has been performed to date. We do not know, for instance, the extent to which this observed racial disparity is because of the preferences of "employers" (prosecutors) or "customers" (family members). Prosecutors could have a preference to file capital homicide charges more often in white victim cases for any number of reasons, including 1) stark racial animus; 2) with white voters in a majority and with finite resources, it may be politically expedient for most prosecutors to charge white victim killings more severely and thereby strategically to play the "tough on crime" card and to appeal to the fear of crime experienced by whites; 3) white prosecutors have greater empathy for white victims and their families because empathy comes easier for individuals like us; and 4) prosecutors think they are more likely to get a death sentence from white jurors when a white rather than a black is killed, and black jurors can be more easily excludable from a capital jury both for legitimate reasons (they automatically oppose death) and illegitimate reasons [prosecutors use peremptory challenges to strike black jurors—see *Batson v. Kentucky* (1986) and *Johnson v. California* (2005)].

The white-race-of-victim effect that we and others before us have observed may also be because of the preferences of family members that prosecutors simply are responding to. It is likely that the family members of white victims more aggressively push prosecutors to seek deaths when their loved ones are killed. African Americans are generally less supportive of the death penalty than whites and are more distrustful of white legal authorities (Huo and Tyler, 2000; Tyler and Huo, 2002). In more often seeking death in white victim cases, then, prosecutors may simply be responding to the demands of their customers. When scholars successfully have documented the story or stories that link victim's race with decisions by prosecutors, an important piece of the causal inference will be in place.