

Rebutting the Myths About Race and the Death Penalty

Kent Scheidegger*

The best models which Baldus was able to devise which account to any significant degree for the major non-racial variables, including strength of the evidence, produce no statistically significant evidence that race plays a part in either [the prosecutor's or the jury's] decisions in the State of Georgia.¹

This is the least-known holding from the best-known case on race and the death penalty, a case that eventually went to the Supreme Court.² It is very strange that this holding is so little known, given the prominence of the Baldus study in debates over race and the death penalty. Just this year, a report of the American Bar Association began its factual discussion of race and the death penalty with what the Baldus study supposedly “showed.”³ Yet the report made no mention at all of the fact that the study had been thoroughly examined in a full trial, with expert testimony on both sides, and found to show nothing of the sort.

This skewed perception is not limited to the Georgia Baldus study. It extends across the field. Any “finding” by a study of any racial “disparity” is trumpeted as proof that the system of capital sentencing is deeply racist, even though it may be the product of flawed methodology, a biased source, or both. Meanwhile, contrary indications from other studies, or sometimes even within the same study, are buried and never brought to the public's attention.

The subject of what these studies show and do not show is a complex one, and a comprehensive treatment is beyond the scope of this short article. The article will trace the development through the principal cases and best-known studies to show that the truth, to the extent we can know it, is quite different from the common perception.

I. *MCGAUTHA* TO *PENRY*: THERE AND BACK AGAIN

In 1971, in *McGautha v. California*,⁴ the U.S. Supreme Court considered an argument that due process required standards for capital sentencing, along the lines

* Legal Director, Criminal Justice Legal Foundation.

¹ *McCleskey v. Zant*, 580 F. Supp. 338, 368 (N.D. Ga. 1984) (emphasis omitted).

² *McCleskey v. Kemp*, 481 U.S. 279 (1987).

³ AMERICAN BAR ASS'N, EVALUATING FAIRNESS AND ACCURACY IN STATE DEATH PENALTY SYSTEMS: THE MISSOURI DEATH PENALTY ASSESSMENT REPORT 332 (2012).

⁴ *McGautha v. California*, 402 U.S. 183 (1971).

suggested by the American Law Institute's draft Model Penal Code. After a characteristically thorough discussion of the history by Justice Harlan, the Court rejected the claim, 6–3. In so doing, the Court rejected the idea that the aggravating and mitigating factors in capital sentencing can be reduced to a defined list.

In light of history, experience, and the present limitations of human knowledge, we find it quite impossible to say that committing to the untrammelled discretion of the jury the power to pronounce life or death in capital cases is offensive to anything in the Constitution. The States are entitled to assume that jurors confronted with the truly awesome responsibility of decreeing death for a fellow human will act with due regard for the consequences of their decision and will consider *a variety of factors*, many of which will have been suggested by the evidence or by the arguments of defense counsel. For a court to attempt to catalog the appropriate factors in this elusive area could inhibit rather than expand the scope of consideration, for *no list of circumstances would ever be really complete*. The *infinite variety of cases and facets to each case* would make general standards either meaningless 'boiler-plate' or a statement of the obvious that no jury would need.⁵

Yet only a year later, the Court made one of the most dramatic flip-flops in its history in *Furman v. Georgia*.⁶ A brief *per curiam* statement (it can hardly be called an opinion) said only "that the imposition and carrying out of the death penalty in these cases constitute cruel and unusual punishment in violation of the Eighth and Fourteenth Amendments."⁷ There was not a word about why. Instead, the reasons why had to be gleaned from the five separate opinions of the justices in the majority. One of the most important laws the states had—the law punishing murder—had been thrown out, and the states had to guess why. For an institution created for the purpose of clarifying the law,⁸ this was an institutional failure of massive proportions.

Why the abrupt reversal? Although racial discrimination is not given as the reason in the *Furman* opinions, it is there between the lines. Among the most perceptive analyses of *Furman* is Justice Thomas's concurring opinion in *Graham v. Collins*.⁹ After tracing the references to racial discrimination in the opinions and the various Justices' conclusions that the case for discrimination in the cases before the Court had not been proved,¹⁰ he concludes, "[i]t cannot be doubted that behind

⁵ *Id.* at 207–08 (emphasis added).

⁶ *Furman v. Georgia*, 408 U.S. 238 (1972).

⁷ *Id.* at 239–40.

⁸ See THE FEDERALIST NO. 78, at 394 (Alexander Hamilton) (Ian Shapiro, ed. 2009).

⁹ *Graham v. Collins*, 506 U.S. 461 (1993) (Thomas, J., concurring).

¹⁰ *Id.* at 479–83.

the Court's condemnation of unguided discretion lay the specter of racial prejudice—the paradigmatic capricious and irrational sentencing factor.”¹¹

Did the justices in the majority in *Furman* think that they did not need to specify how to fix the statutes because there would be insufficient interest in reinstating capital punishment, and the practice would die? If so, they were very much mistaken. By 1976, restoration legislation had been passed by Congress and the legislatures of thirty-five states.¹² Public support for the death penalty, which had dipped below a majority a few years before *Furman*, shot upward in the years following, reaching a peak of 80% by 1994.¹³

Yet all these legislative bodies had to guess what was constitutional, due to the failure of the Court to specify. The legislatures with the largest and most sophisticated resources—Congress, California, and New York—all believed that mandatory sentencing was required. Making the sentence apply mechanically based on a few, objective circumstances would eliminate the potential for discrimination that was the real basis of *Furman*. It was a well-founded belief,¹⁴ but the Court would soon declare forbidden what it previously implied was required.¹⁵

Georgia made the least change in its pre-*Furman* statute. Eligibility for the death penalty was narrowed to a subset of murders by a finding of aggravating circumstances, but from that point those circumstances had no special role, and the jury can consider all relevant circumstances in reaching its final decision. To the surprise of many observers, this system—only one step removed from the ones struck down in *Furman*—drew praise from the Court.¹⁶

Texas sought to meet the structure requirement by making the sentencing decision turn on the answers to specific questions, while Florida adopted a system quite similar to the Model Penal Code draft, directing the jury to weigh. The Supreme Court initially approved both systems with no hint that juries needed to be instructed on any factors other than those in the approved statutes,¹⁷ but the Court would eventually betray both states and effectively strike down what it had previously approved.¹⁸ In *Lockett v. Ohio*, a plurality held that states are not only

¹¹ *Id.* at 482.

¹² See *Gregg v. Georgia*, 428 U.S. 153, 179–80 (1976) (lead opinion). For *Gregg* and its companion cases, the joint opinions of Justices Stewart, Powell, and Stevens are cited in this article as the “lead opinion.” These are the opinions that announced the judgment of the Court and have been treated by the Court as embodying the holdings of the cases.

¹³ Jeffrey M. Jones, *Support for the Death Penalty 30 Years After the Supreme Court Ruling*, GALLUP NEWS SERVICE, June 30, 2006, <http://www.gallup.com/poll/23548/Support-Death-Penalty-Years-After-Supreme-Court-Ruling.aspx>.

¹⁴ See *Rockwell v. Superior Court*, 556 P.2d 1101, 1117 (Cal. 1976) (Clark, J., concurring).

¹⁵ See *Roberts v. Louisiana*, 428 U.S. 325, 356 (1976) (White, J., dissenting).

¹⁶ *Gregg*, 428 U.S. at 203–04; see also *Zant v. Stephens*, 462 U.S. 862, 878–79 (1983).

¹⁷ *Jurek v. Texas*, 428 U.S. 262, 268 (1976) (lead opinion); *Proffitt v. Florida*, 428 U.S. 242, 247 (1976) (lead opinion).

¹⁸ See Kent S. Scheidegger, *Capital Punishment in 1987: The Puzzle Nears Completion*, 15

permitted to let the jury consider “any aspect of a defendant’s character or record and any of the circumstances of the offense that the defendant proffers as a basis for a sentence less than death,” they are required to do so.¹⁹ That holding has been reaffirmed and amplified in multiple decisions since.²⁰

Did the Supreme Court in *Lockett* and its progeny wipe out whatever antidiscriminatory good it had achieved with *Furman*? Justices White, Scalia, and Thomas thought so.²¹ In any event, for better or worse, an individual weighing of all relevant factors is not a state’s choice, it is the way a state is constitutionally required to operate its capital sentencing system. A mathematical model of capital sentencing that fails to weigh or leaves out relevant factors is simply wrong. That is one of many troubles with mathematical models of human behavior.

II. THE TROUBLES WITH MODELS

Mathematical models have a prominent role in the debate over race and the death penalty. Because most lawyers and judges know so little about them, I will give a very brief and simplified introduction here, making no pretense of being an expert on the subject.

“All models are wrong but some are useful” is a saying in the social sciences, attributed to statistician George Box.²² In physics, mathematical models can represent reality exactly. Pass a one milliamp direct current through a 1000 ohm resistor, and the voltage across it is given exactly by the simple formula $V = IR$, one volt. People are not so easily or so reliably modeled. We are less numerous, more complex, and more varied than the electrons in the previous example. In modeling human behavior, “we are trying to force the ugly stepsister’s foot into Cinderella’s pretty glass slipper. It doesn’t fit without cutting off some essential parts.”²³ The wisdom in Box’s maxim is that models can be useful for some purposes if we are constantly aware of their limitations. Forgetting those limitations can be disastrous. Too much faith in models was a contributing factor in the 2008 subprime loan meltdown.²⁴

A very common misuse of studies is to take a result that says two things are correlated (A tends to go with B) and jump to the conclusion that A causes B. The

W. ST. U. L. REV. 95, 115 (1987).

¹⁹ *Lockett v. Ohio*, 438 U.S. 586, 604 (1978) (plurality opinion).

²⁰ See *Eddings v. Oklahoma*, 455 U.S. 104, 113–16 (1982); *Penry v. Lynaugh*, 492 U.S. 302, 328 (1989).

²¹ See *Lockett*, 438 U.S. at 622 (White, J., concurring); *Walton v. Arizona*, 497 U.S. 639, 662 (1990) (Scalia, J., concurring); *Graham v. Collins*, 506 U.S. at 493–94 (Thomas, J., concurring).

²² Julian J. Faraway, Practical Regression and Anova using R, 47 (July 2002) (unpublished manuscript) (on file at The Comprehensive R Archive Network), available at cran.r-project.org/doc/contrib/Faraway-PRA.pdf.

²³ EMANUEL DERMAN, MODELS.BEHAVING.BADLY: WHY CONFUSING ILLUSION WITH REALITY CAN LEAD TO DISASTER, ON WALL STREET AND IN LIFE 196 (2011).

²⁴ See *id.* at 150.

classic, albeit apocryphal, example is a study showing a correlation between sales of ice cream and crimes of violence.²⁵ This proves that ice cream causes crime, and therefore ice cream should be banned. This obviously false conclusion is the result of the kind of error that is regularly committed in public policy discussions when the falsity of the conclusion is less obvious.²⁶

There are other reasons two numbers might go up and down together. One possibility is that A and B can be correlated if both are caused by C, in this example, hot weather. When the temperature climbs, people who like ice cream tend to buy more ice cream, and people prone to violence tend to commit more acts of violence. Banning ice cream would therefore do nothing to help the problem. It might make it worse.

Multiple regression is a technique used to explain or model the relationship between a variable, called the dependent variable, and multiple other variables, called independent variables.²⁷ If a relationship is found, it is called an "effect." That is an unfortunate term because it implies to the uninitiated that a cause-and-effect relationship has been found. Regression cannot tell us that.

Regression begins by building a mathematical model. The simplest kind would plot on a graph as a straight line and so is called a linear model. So for the ice cream example we might hypothesize:

$$C = A_1 + A_2 * I + A_3 * T + e$$

where C is the crime rate on a given day, I is sales of ice cream on that day, and T is the peak daily temperature. The As are constant across all days, to be determined in the regression procedure. The e is the error term, the difference between each value of C and the predicted value from combining the other terms. It includes all the factors going into the crime rate, which we hope are random with respect to the variables we are interested in.

We load the data into the regression computer program, and it gives us estimates for the coefficients (the As) and some statistics to help us decide if we can have confidence in the result. If all goes well, the results tell us that the daily variation in crime is actually related to temperature, and the ice cream has no independent relation to crime. In other words, there is no correlation between ice cream and crime once we have controlled for temperature.

What can go wrong? Many things.²⁸ Some are too technical to explain here. One problem is an insufficient number of data points to be confident the relation

²⁵ For one real, but tongue-in-cheek, study along these lines, see Eugene Volokh, *Academic study reveals*., THE VOLOKH CONSPIRACY (July 13, 2004) www.volokh.com/2004/07/13/academic-study-reveals/.

²⁶ See generally Arnold Barnett, *How Numbers Can Trick You*, TECH. REV., Oct. 1994, at 38, 39 (misuse of statistics in public policy debates).

²⁷ See Faraway, *supra* note 22, at 13.

²⁸ See *id.* at 46–47.

we observe is not pure chance. There is a test for that, and the rule of thumb is that we declare the result "statistically significant" if the chance of the observed relation happening at random is less than five percent. That is only a rule of thumb, not Revealed Truth as it is too often treated.²⁹

The data may be suspect. Some of it may be just plain wrong. Sometimes there are missing data for some points in the study. More fundamentally, the concept we are really interested in may not be a simple number or a yes/no, as the technique demands. Is peak daily temperature really the measure of how hot a day is for the purpose of flaring tempers? What about humidity, breeze, or the extent to which it cools off in the evening?

We know our model does not include everything that goes into the result. What if the omitted factors predominate, so that the factors in our model explain very little of the result? Even worse, what if an omitted factor is related to the one we are trying to measure? Suppose in the above example we left out temperature but included other factors, say weekend/weekday, phase of the moon, and whether a locally important sporting event was on. Then the spurious relation between ice cream and crime might well reappear. Can we just throw in every factor conceivably relevant and build a model with a huge number of factors? That creates problems of its own.³⁰

The simple point here is that mathematical modeling is not magic. Just because the numbers come out of the computer with lots of statistics and graphs does not mean we can depend on the conclusion being true. A lot can go wrong, and studies need to be challenged by someone with the incentive and expertise to do so.

III. THE *MCCLESKEY* CASE

The best known case on race and the death penalty is the case of Warren McCleskey,³¹ a habitual criminal³² who shot and killed police officer Frank Schlatt in the course of robbing a furniture store in Atlanta, Georgia.³³ On federal habeas corpus review, McCleskey claimed that his sentence was tainted by racial discrimination. The principal evidence for this claim was a pair of studies by David Baldus and others, the second one commissioned by the NAACP Legal Defense and Education Fund for the specific purpose of attacking the death

²⁹ See Jacob Cohen, *Things I Have Learned (So Far)*, 45 AM. PSYCHOLOGIST 1304, 1304 (1990).

³⁰ See *id.* at 1304-05.

³¹ His name is spelled McClesky in the earlier cases and McCleskey in the later ones, including the federal cases discussed in this article.

³² See *McCleskey v. Kemp*, 753 F.2d 877, 882 (11th Cir. 1985), *aff'd*, 481 U.S. 279 (1987) (three prior armed robbery convictions).

³³ *McCleskey v. Zant*, 580 F. Supp. 338, 345 (N.D. Ga. 1984), *aff'd in part, McCleskey*, 753 F.2d at 885.

penalty.³⁴ The case is best known for the Supreme Court's holding that *even if* the Baldus study showed what it claimed to show, McCleskey did not have a case.³⁵ This article will begin with the important, but largely forgotten, findings of the District Court.

A. *The Trial*

The study used multiple regression analysis, discussed in the previous section, a technique then "relatively new to the law."³⁶ Judge Owen Forrester heard experts on both sides: Baldus, George Woodworth, and Richard Berk for McCleskey and Joseph Katz and Roger Burford for the state.³⁷

Judge Forrester noted, "no statistical analysis, much less a multivariate analysis, is any better than the accuracy of the data base."³⁸ A more common way of saying this in the computer business is "garbage in, garbage out." To be fair to the researchers, extracting reliable data on the many factors that go into a capital sentencing decision from the case files is a huge task, perhaps an impossible one. But we are concerned with the quality of the product, not the quality of the effort. After a thorough review, Judge Forrester concluded that "the data base has substantial flaws and . . . petitioner has failed to establish by a preponderance of the evidence that it is essentially trustworthy."³⁹

Even assuming the validity of the data base, Judge Forrester found further problems. The statistical test for showing how much of the variation in the result is accounted for by the model found that only half was accounted for, and the rest was unexplained. "None of the models presented have accounted for the alternative hypothesis that the race effects observed cannot be explained by unaccounted-for factors."⁴⁰

Another problem, difficult for nonexperts to understand, is the problem of multicollinearity. Suffice it to say that when two variables in the equation are related to each other, the regression technique is not very good at separating out the effect of one versus the effect of the other. Katz's analysis of the Georgia data show that victim race is strongly correlated with legitimate sentencing variables and offender-victim race combinations are even more so. Killings during armed robberies were 33.3% of the white-victim cases and only 7.4% of the black-victim

³⁴ See DAVID C. BALDUS, GEORGE G. WOODWORTH & CHARLES A. PULASKI, JR., *EQUAL JUSTICE AND THE DEATH PENALTY* 44 (1990).

³⁵ *McCleskey v. Kemp*, 481 U.S. 279, 291, n.7 (1987) (assuming, not deciding, validity of the study); *id.* at 298–99 (study does not establish an equal protection claim); *id.* at 312–13 (study does not establish Eighth Amendment claim).

³⁶ *McCleskey*, 580 F. Supp. at 350.

³⁷ *Id.* at 352–53.

³⁸ *Id.* at 354.

³⁹ *Id.* at 360.

⁴⁰ *Id.* at 362.

cases. The killer was a stranger to the victim in 35.8% of white-victim cases but only 18.8% of black-victim cases. Looking at black-perpetrator, white-victim cases, robberies are a staggering 67.1% and stranger-murders are 70.6%.⁴¹ Crimes of predation, where the victim is chosen simply because he has something the perpetrator wants, strike particular fear into people's hearts. "That could have been me." This is an entirely legitimate factor, strongly correlated with race, and multicollinearity limits the ability of regression analysis to account for it.

Finally, even overlooking all the foregoing problems and putting Baldus's results into a table, Judge Forrester observed:

The coefficients produced by the 230-variable model on the Charging and Sentencing Study data base produce no statistically significant race of the victim effect either in the prosecutor's decision to seek the death penalty or in the jury sentencing decision. A 200-variable model based on the Procedural Reform data base shows a statistically significant race of the victim effect at work on the prosecutor's decision-making, but that model is totally invalid for it contains no variable for strength of the evidence, a factor which has universally been accepted as one which plays a large part in influencing decisions by prosecutors. Neither model produces a statistically significant race of the defendant effect at the level where the prosecutor is trying to decide if the case should be advanced to a penalty trial. Neither model produces any evidence that race of the victim or race of the defendant has any statistically significant effect on the jury's decision to impose the death penalty. The significance of this table cannot be overlooked. The death penalty cannot be imposed unless the prosecutor asks for a penalty trial and the jury imposes it. *The best models which Baldus was able to devise which account to any significant degree for the major non-racial variables, including strength of the evidence, produce no statistically significant evidence that race plays a part in either of those decisions in the State of Georgia.*⁴²

That is a dense paragraph, so let's unpack it. The most important finding is buried in the middle. Accepting for the sake of argument the death penalty opponents' best known study on its own terms, there is no evidence of discrimination against black defendants, but rather a refutation of any such claim. The primary concern underlying the *Furman* decision⁴³ has been refuted in the post-*Gregg* era. This is wonderful news and cause for celebration. Yet despite

⁴¹ See Joseph L. Katz, Warren McCleskey v. Ralph Kemp: *Is the Death Penalty in Georgia Racially Biased?*, in CAPITAL PUNISHMENT: A BALANCED EXAMINATION 403-06 (Evan J. Mandery ed., 2005); see also *McCleskey*, 580 F. Supp. at 363-64.

⁴² *McCleskey*, 580 F. Supp. at 367-68.

⁴³ See *Furman v. Georgia*, 408 U.S. 238, 309 (1972) (Stewart, J., concurring); see also *supra* text accompanying note 6.

replication of this result in numerous other studies, some of which are described below, the result remains nearly unknown to the general public. A presidential candidate in 2012 declared in a televised debate that it is “very clear” black defendants “suffer the consequences of the death penalty disproportionately.”⁴⁴ Nonsense. The opponents’ own studies say just the opposite.

Race of victim “effects” are “statistically significant” at one point in the process but not others. This does not mean that racial animus on the part of the decision-maker is the reason for the observed “effect.” Another possibility is a legitimate factor, correlated with race but not accounted for in the model. Judge Forrester notes a whopper of a factor: strength of the evidence of guilt. Contrary to the prejudiced image popular in some circles, capital case prosecutors are typically very much concerned with the justice of the case and the possibility, however remote, of executing an innocent person. I know this from a quarter-century of experience working with capital case prosecutors all over the country. Residual doubt of guilt is a powerful and entirely proper reason not to seek the death penalty, even if the crime is especially heinous.

Is there reason to believe that strength of the case is correlated with race of the victim? Regrettably, yes. Building a case depends on the willingness of witnesses to come forward and the credibility of those witnesses. Community trust in the police, social norms against “snitching,” fear of reprisal, and the likelihood that a witness has a criminal record of his own are all likely to be correlated with race.

After further discussion of the limitations of regression and other topics, Judge Forrester concluded that McCleskey had failed to make his case. Not only did he fail to prove discrimination in his individual case, the point the Supreme Court would later hold dispositive, Baldus et al. failed to prove (and the State’s experts succeeded in rebutting) the basic claims made in the Baldus study.⁴⁵ They did not just fail; they failed dismally. The Baldus study lay in shreds when Judge Forrester got through with it.

Yet the Baldus study “has received an undeservedly good press.”⁴⁶ When Baldus died last year, a story in the New York Times praised his work, calling it “meticulous,” never mentioning the thorough, careful judicial finding that it utterly failed to prove what Baldus claimed.⁴⁷ The most astonishing misstatement of the status of the Baldus study came from the Supreme Court itself. In his notorious dissent from denial of certiorari in the capital case of *Callins v. Collins*, Justice Harry Blackmun made the jaw-dropping assertions that the Baldus study is “highly

⁴⁴ Kent Scheidegger, *Disproportionate to What?*, CRIME AND CONSEQUENCES BLOG (Jan. 17, 2012, 9:04 AM), <http://www.crimeandconsequences.com/crimblog/2012/01/disproportionate-to-what.html>.

⁴⁵ 580 F. Supp. at 379–80.

⁴⁶ *Statistical Evidence of racial discrimination in the death penalty: Statement to the N.C. H. Select Comm. on Capital Punishment*, 2005–2006 Sess. (N.C. 2006) (statement of Elliot Cramer) [hereinafter *Cramer Statement*], available at <http://ourpaws.info/cramer/death/talk.txt>.

⁴⁷ Adam Liptak, *David C. Baldus, 75, Dies; Studied Race and the Law*, N.Y. TIMES, June 15, 2011, at B13.

reliable” and “as far as I know, there has been no serious effort to impeach the Baldus study.”⁴⁸ Did Justice Blackmun not read the district court opinion in a case where he wrote a dissent arguing for reversal of its judgment? It is hard to come to any other conclusion.

B. *The Appeal*

One reason that the Baldus study has “received an undeservedly good press” derives from the unusual way the *McCleskey* case was handled on appeal. The Court of Appeals for the Eleventh Circuit, sitting en banc, commended the district court “for its outstanding endeavor” in analyzing the validity of the Baldus study, and there is little doubt that a review of the factual finding that the study was invalid would have been affirmed under the applicable “clearly erroneous” standard.⁴⁹ However, the appellate court decided not to review that aspect of the decision, but instead decided to proceed “by *assuming* the validity of the study and rest[ing their] holding on the decision that the study, even if valid, not only supports the district judge’s decision under the clearly erroneous standard of review, but compels it.”⁵⁰

Assuming disputed facts in a party’s favor is standard practice when the other party seeks a judgment without a trial, such as on a motion for summary judgment.⁵¹ Making such an assumption on appeal following a full trial and determination of those facts is much more rare.

On certiorari, the Supreme Court reviewed only the court of appeals’s judgment on this basis. It did not go back and review the district court decision on validity.⁵² This is clear enough for those who read the opinion with any degree of care, and the frequent miscitation of the opinion as accepting or endorsing the Baldus study is the fault of those who cite it that way, rather than the opinion itself.⁵³

Unfortunately, the Supreme Court opinion, in the course of summarizing Baldus’s claimed results, committed a gross error that has contributed to an extreme and erroneous perception ever since. The brief for *McCleskey* by John Charles Boger misleadingly stated, “Professor Baldus testified that his best statistical model . . . revealed that after taking into account most legitimate reasons

⁴⁸ *Callins v. Collins*, 510 U.S. 1141, 1153–54 (1994).

⁴⁹ *See McCleskey v. Kemp*, 753 F.2d 877, 894–95 (11th Cir. 1985).

⁵⁰ *Id.* at 895 (emphasis added).

⁵¹ *See, e.g., Saucier v. Katz*, 533 U.S. 194, 201 (2001).

⁵² *See McCleskey v. Kemp*, 481 U.S. 279, 291 n.7 (1987).

⁵³ For example, at a mock trial event in London in March 2010, the prosecutor asked me repeatedly if the Supreme Court had “accepted” the Baldus study and seemed genuinely surprised at my answer it had not. *See Video: Defending American Justice*, testimony of Kent Scheidegger (Criminal Justice Legal Foundation 2010), available at <http://www.cjlf.org/media/dpontrial/London.htm>.

for sentencing distinctions, the *odds* of receiving a death sentence were still more than 4.3 times greater for those whose victims were white than for those whose victims were black.”⁵⁴ The opinion of the Court says, “defendants charged with killing white victims were 4.3 *times as likely* to receive a death sentence as defendants charged with killing blacks.”⁵⁵ To one unschooled in statistics, these statements may seem equivalent. They are not. This is literally a textbook example of how to lie with statistics.

Arnold Barnett, a professor at MIT, wanted to illustrate the misuse of statistics in the popular press with the goal “to discourage fellow citizens from taking a strong position or course of action based solely on a press report.”⁵⁶ The misleading “odds ratio” of *McCleskey* stuck out as a prime example. Quoting the passage above from the opinion and a similar one from the New York Times, Barnett notes, “the Supreme Court, the New York Times, and countless other newspapers and commentators were laboring under a major misconception.”⁵⁷ He then went through the math to show how an “odds ratio” can wildly exaggerate the degree of disparity. He concludes, “[b]lame for the confusion should presumably be shared by the judges and the journalists who made the mistake and the researchers who did too little to prevent it.”⁵⁸ He left out the lawyers.

Regrettably, Barnett’s exposure of this misleading use of statistics has had no discernible effect. To this day, death penalty opponents seeking to play the race card, including Boger himself, continue to exploit the confusion to inflate their claims of disparity.⁵⁹

The legal holding of the Supreme Court’s *McCleskey* decision has been dissected many times by many commentators, so I will not belabor that point here. The decision was a major legal victory for the state, effectively shutting down these kinds of statistical claims in federal courts. Yet the coverage of the decision handed an undeserved public relations victory to the opponents, creating a false public impression that the case of “race of victim bias” had been proved and that the degree of disparity is much greater than the study even claimed it was.

IV. NEW JERSEY

While the *McCleskey* decision shut down litigation of this type in federal courts, state courts were not necessarily bound to the result. They can interpret their own state’s constitution and laws to extend protection to defendants that federal law does not, and the U. S. Supreme Court will not review these decisions

⁵⁴ Brief for Petitioner at 15–16, *McCleskey v. Kemp*, No. 84-6811 (1986), 1986 U.S. S. Ct. Briefs LEXIS 489 (emphasis added).

⁵⁵ 481 U.S. at 287 (emphasis added).

⁵⁶ Barnett, *supra* note 26, at 38–39.

⁵⁷ *Id.* at 43.

⁵⁸ *Id.*

⁵⁹ See *Cramer Statement*, *supra* note 46.

based on independent state grounds.⁶⁰

The New Jersey Supreme Court, prior to the repeal of the death penalty in that state, was an exceptionally friendly forum for capital defendants.⁶¹ That court rejected *McCleskey* on independent state grounds.⁶² It appointed a special master to examine the matter, curiously choosing none other than David Baldus,⁶³ an advocate for one side of the debate and the author of the study so severely criticized by the District Court in *McCleskey*. Early analyses were hampered by an inadequate number of cases, however, and the results were inconclusive.⁶⁴ After the data base grew with the addition of new cases, the court appointed a new special master, Judge Richard Cohen.⁶⁵ “On the statistical evidence before him, Judge Cohen concluded that he did not find ‘relentless documentation or even a preponderance in the direction of the existence of any race bias.’”⁶⁶ The court was determined to continue reviewing this issue, and it appointed a third special master, Judge David Baime.⁶⁷

Judge Baime remained the special master for several years and produced a series of annual reports. In these reports, the available data were analyzed with three different methods.⁶⁸ The result of the analysis was the statistical evidence did not support a claim of bias on either the race of the defendant or the race of the victim.

The statistical evidence does not support the thesis that the race of the victim affects the likelihood that the defendant will receive the death penalty. We add that the available statistical evidence discloses that African-American defendants who kill White victims are no more likely to receive the death penalty than African-American defendants who kill African-American victims.⁶⁹

In a critically important finding, Judge Baime noted that a disparity had appeared initially, but it turned out that race of the victim was confounded with jurisdiction. Fewer black-victim cases proceed to penalty trial because the

⁶⁰ See, e.g., *Michigan v. Long*, 463 U.S. 1032, 1040–41 (1983).

⁶¹ See Kent Scheidegger, *Statement Before the New Jersey Death Penalty Study Commission*, CRIM. JUST. LEGAL FOUND. 1–2 (2006), <http://www.cjlf.org/files/NJDPTestimony.pdf>.

⁶² See *State v. Marshall*, 613 A.2d 1059, 1108–09 (N.J. 1992).

⁶³ See *id.* at 1063.

⁶⁴ See *State v. Loftin*, 724 A.2d 129, 152 (N.J. 1999).

⁶⁵ See *id.* at 153–54.

⁶⁶ *Id.* at 160.

⁶⁷ See *id.* at 232.

⁶⁸ DAVID S. BAIME, REPORT TO THE SUPREME COURT SYSTEMIC PROPORTIONALITY REVIEW PROJECT: 2000–2001 TERM 1 (2001).

⁶⁹ *Id.* at 61.

counties in which most black people live take fewer of their cases to penalty trial.⁷⁰ “New Jersey is a small and densely populated state. It is, nevertheless, a heterogeneous one. It is thus not remarkable that the counties do not march in lockstep in the manner in which death-eligible cases are prosecuted.”⁷¹

Thus, the two studies fully adjudicated in court have come to consistent results. There is no race-of-defendant bias. A connection that initially appears between race of the victim and death sentencing rates disappears when further legitimate factors are considered.

V. UNADJUDICATED STUDIES

The other studies in this area have not been the subject of completed adjudication. Greater caution is therefore in order when considering their conclusions. As we saw in the *McCleskey* case, the author of a study can claim his study supports a particular conclusion, but upon a challenge and adjudication, the result may be just the opposite.

A. Maryland

“Large Racial Disparity Found Study of Md. Death Penalty,” read a front-page headline in the Washington Post in 2003.⁷² Given America’s obsession with race, there was no shortage of people willing to jump to the conclusion that this study once again confirmed racist discrimination against black people in the administration of criminal justice. A closer look at the study⁷³ reveals a more nuanced picture.

The study was a large and detailed one, and to the researchers’ credit they made a strong attempt to capture and quantify the legitimate factors that go into a prosecutor’s decision to seek the death penalty and the jury’s decision to impose it.⁷⁴ Some factors, however, will always defy quantification. Strength of the evidence, the factor the *McCleskey* court noted was omitted from Baldus’s main model,⁷⁵ is sought to be quantified with such factors as whether and how many eyewitnesses testify and whether there is physical evidence linking the defendant to the crime. Such “yes or no” or “how many” questions cannot begin to capture this critical variable, though. Physical evidence may be anything from a very weak link (e.g., a hair found at a scene where defendant admits he was for innocent

⁷⁰ See *id.* at 61–62.

⁷¹ *Id.* at 62.

⁷² Susan Levine & Lori Montgomery, *Large Racial Disparity Found Study of Md. Death Penalty*, WASH. POST, Jan. 8, 2003, at A1.

⁷³ R. PATERNOSTER ET AL., AN EMPIRICAL ANALYSIS OF MARYLAND’S DEATH SENTENCING SYSTEM WITH RESPECT TO THE INFLUENCE OF RACE AND LEGAL JURISDICTION (2003).

⁷⁴ See *id.* at tbl.9 (listing 123 covariates).

⁷⁵ *McCleskey v. Zant*, 580 F. Supp. 338, 367–68 (N.D. Ga. 1984).

reasons a week earlier) to virtually conclusive (e.g., a DNA match to the victim's blood on the defendant's blood-soaked shirt). Eyewitnesses vary widely in credibility, for a host of reasons. That is why we have safeguards such as the right to confront and cross-examine witnesses and the right to counsel at critical pretrial stages as well as at trial.⁷⁶

The results section of the study begins with a description of race and county patterns without adjusting for case characteristics.⁷⁷ These results are utterly irrelevant. We know from past research that crime characteristics that legitimately go into sentencing decisions do correlate with race,⁷⁸ so "disparities" in unadjusted data tell us nothing of policy significance.

In the adjusted analysis, the study concluded, "*there is no evidence that the race of the defendant matters at any stage once case characteristics are controlled for.*"⁷⁹ That should have been the headline in the next day's newspapers. It is the clearest outcome in the study and the one most salient to the question of whether criminal defendants are being treated unfairly on the basis of race. Instead, this result got little publicity.

Another clear result is that Maryland's locally elected prosecutors seek the death penalty at different rates.⁸⁰ Of course they do. That is why we elect them locally in almost all states, to give the people of the community a voice in how vigorously the criminal law will be enforced in their community. If the people of downtown Baltimore City elect a prosecutor who seeks the death penalty relatively rarely while the people of suburban Baltimore County elect one who seeks it more often, that is local democracy working as designed.

When it comes to race of the victim, the picture is more muddled. Recall that in New Jersey the special master found that jurisdiction was confounded with race.⁸¹ It is in Maryland as well. This should not surprise anyone. In the post-*Gregg* era, the Gallup Poll has found that the death penalty has been supported by two-thirds to three-quarters of whites while support among blacks is much lower at two-fifths to a little over half.⁸² Naturally, a jurisdiction with a large black population elects a prosecutor who seeks the death penalty more selectively, and those are the jurisdictions where most of the black-victim homicides occur.

Do black-victim cases remain less likely to result in a death sentence after controlling for both case characteristics and jurisdiction? The Paternoster study gives us a mixed result. Using one method, the "effect" remains "statistically

⁷⁶ See *Perry v. New Hampshire*, 132 S. Ct. 716, 728–29 (2012).

⁷⁷ PATERNOSTER ET AL., *supra* note 73, at 20.

⁷⁸ See Katz, *supra* note 41 and accompanying text.

⁷⁹ PATERNOSTER ET AL., *supra* note 73, at 31 (emphasis in original).

⁸⁰ See *id.* at 28–31.

⁸¹ See BAIME, *supra* note 68 and accompanying text.

⁸² Lydia Saad, *Racial Disagreement Over Death Penalty Has Varied Historically*, GALLUP (July 30, 2007), <http://www.gallup.com/poll/28243/Racial-Disagreement-Over-Death-Penalty-Has-Varied-Historically.aspx>.

significant,” meaning it passes the traditional rule of thumb for saying we are confident the observed connection is not due to pure chance. Using another method, it is not.⁸³

This is hardly compelling evidence that race of the victim *as such* is the reason for observed differences. It may or may not pass the standard statistical criterion for something other than pure chance, depending on which model is used. But on top of pure chance, there are all the other reasons why these kinds of statistical studies may not reflect reality. Among other issues, there are the limited ability of the studies to quantify the case characteristics noted above and the possibility that other factors correlated with race and not considered at all in the study may influence the result. For example, do prosecutors in Maryland consider the wishes of the victim’s family in making their decision? That factor is likely correlated with race.⁸⁴ The study acknowledges the existence of this factor⁸⁵ but does not control for it.⁸⁶

If, with all the effort that went into this study, the best they can do is show an effect on the ragged edge of the most basic criterion, the logical conclusion is that race of the victim is probably *not* a major factor in deciding who is sentenced to death. That decision is primarily the result of legal criteria, the circumstances of the case, and the democratic choice of the people of the local jurisdiction, as it should be. We can say that on the face of the study even without the kind of adversarial testing we had in the *McCleskey* case.

Although the study has not been the subject of an adversary proceeding, it has been challenged in a published article.⁸⁷ The lead author is Richard Berk, an expert witness for the *defense* in the *McCleskey* case. This is significant, because many academics who do research on the death penalty reliably produce results that favor one side, raising a suspicion of partisan bias.

The Berk, Li, and Hickman paper is technically dense and difficult for anyone other than a statistician to understand. Suffice it to say that they take the same data analyzed by Paternoster et al., apply different techniques, and get different results. “For both capital charges and death sentences, race either played no role or a small one that is very difficult to specify. In short, it is very difficult to find convincing evidence for racial effects in the Maryland data”⁸⁸ Their point is not to support the death penalty but rather to point out that results from this kind of modeling are “fragile.”⁸⁹ What “studies show” is not necessarily so. Results are

⁸³ PATERNOSTER ET AL., *supra* note 73, at 33.

⁸⁴ Saad, *supra* note 82 and accompanying text.

⁸⁵ See PATERNOSTER ET AL., *supra* note 73, at 14–15.

⁸⁶ See *id.*, tbl. 9 (factor not listed).

⁸⁷ See Richard Berk, Azusa Li & Laura J. Hickman, *Statistical Difficulties in Determining the Role of Race in Capital Cases: A Re-analysis of Data from the State of Maryland*, 21 J. QUANTITATIVE CRIMINOLOGY 365, 367–68 (2005).

⁸⁸ *Id.* at 386.

⁸⁹ *Id.*

heavily dependent on choices made in the modeling process.

B. *The Federal System*

One more set of studies warrants particular mention because it is an exceptionally egregious example of how the popular perception diverges from the reality. On September 11, 2000, officials at the U.S. Department of Justice were talking to the press about a report to be released the next day.

In the first comprehensive review of the federal death penalty since it was reinstated in 1988, the Justice Department has found significant racial and geographic disparities, say officials who have seen the report.

In 75 percent of the cases in which a federal prosecutor sought the death penalty in the last five years, the defendant has been a member of a minority group, and in more than half the cases, an African-American, according to the report, which officials said the Justice Department would release on Tuesday.

"It's troubling," said an administration official who has reviewed the data. "The president has expressed concern about the problem, and this backs that up." Another administration official described the report as "disturbing."⁹⁰

This is an extraordinarily odd way for the release of a government report to be handled. Officials making statements about a report before the report is released virtually guarantees that the news coverage will be based on the statements and not on the report itself. That is especially true with inflammatory statements by the officials that border on accusing their own department of racism.

The report actually released the next day,⁹¹ however, was not a "comprehensive review." It was a compilation of raw race statistics without any adjustment for case characteristics. As discussed earlier, such unadjusted data is essentially meaningless. Further, the figures showing large percentages of minorities were presented without the context of the pool of cases subject to federal capital prosecution. Contrary to the press statements of the anonymous officials, the data are not "disturbing" if one is familiar with these facts, and they do not, by themselves, back anything up. Releasing unadjusted data, knowing how it was likely to be misinterpreted and how inflammatory the issue was, was questionable at best. The anonymous commenting campaign the day before was

⁹⁰ Raymond Bonner & Marc Lacey, *Pervasive Disparities Found in the Federal Death Penalty*, N.Y. TIMES, Sept. 12, 2000, at A1.

⁹¹ U.S. DEP'T OF JUSTICE, *THE FEDERAL DEATH PENALTY SYSTEM: A STATISTICAL SURVEY (1988-2000)* (2000).

grossly irresponsible conduct.

The first step toward a responsible assessment was to establish the relevant pool of cases for federal capital prosecution. This task was completed ten months later. Murder, as such, is not a federal offense in most of the country, and federal death penalty cases largely involve killings in the course of drug trafficking operations. “In areas where large-scale, organized drug trafficking is largely carried out by gangs whose membership is drawn from minority groups, the active federal role in investigating and prosecuting these crimes results in a high proportion of minority defendants in federal cases, including a high proportion of minority defendants in potential capital cases arising from the lethal violence associated with the drug trade.”⁹² The supposedly “disturbing” raw numbers in the initial report reflected the demographic realities of drug trafficking in America in the 1990s and the fact that this segment of homicides represents most of the federal capital cases. Few people familiar with the operations of the U.S. Department of Justice would have needed a study to tell them that.

Of course a comparison of capital cases with the pool of potentially capital cases is only a first step. The next step is to attempt to control for case characteristics, as in the studies previously described. At this point, the National Institute of Justice did something unique. The data base was created in a manner similar to the other studies, but then the data were given to three independent research teams who analyzed it separately with their own methods. Only after their independent analyses were complete did they get together to compare results.⁹³

All three teams found the results along the same lines as those found in the *McCleskey* court’s assessment of the Baldus study, in the New Jersey special master’s report, in Berk and Li’s reanalysis of the Maryland data, and at least partially in the Paternoster analysis of the Maryland data. “When we look at the raw data and make no adjustment for case characteristics, we find the large race effects noted previously—namely, a decision to seek the death penalty is more likely to occur when the defendants are white and when the victims are white. However, these disparities disappear when the data coded from the AG’s case files are used to adjust for the heinousness of the crime.”⁹⁴ This result gains considerable credibility by the convergence of three independent teams. “Nevertheless, the three teams agreed that their analytic methods cannot provide definitive answers about race effects in death-penalty cases. Analyses of observational data can support a thesis and may be useful for that purpose, but such analyses can seldom prove or disprove causation.”⁹⁵

⁹² U.S. DEP’T OF JUSTICE, *THE FEDERAL DEATH PENALTY SYSTEM: SUPPLEMENTARY DATA, ANALYSIS AND REVISED PROTOCOLS FOR CAPITAL CASE REVIEW* para. 4 (2001).

⁹³ STEPHEN P. KLEIN, RICHARD A. BERK & LAURA J. HICKMAN, RAND, *RACE AND THE DECISION TO SEEK THE DEATH PENALTY IN FEDERAL CASES* xv–xvi (2006), *available at* http://www.rand.org/content/dam/rand/pubs/technical_reports/2006/RAND_TR389.pdf.

⁹⁴ *Id.* at xvii.

⁹⁵ *Id.* at xx.

VI. SO WHAT DOES IT ALL MEAN?

What useful conclusions can we draw from this confusing mix of complex, sometimes flawed, sometimes conflicting studies? First, we must know what we do not know. The caution voiced in the last study quoted above must be taken to heart. These studies will never provide definitive proof.

Second, the most “robust” result, the one that comes up again and again, study after study, jurisdiction after jurisdiction, is the absence of any significant evidence of racial bias against minority defendants. Far and away the most disturbing result, if it occurred, would be an indication that people are on death row who would not be there if they were a different race. The absence of such a result is a remarkable achievement, worthy of celebration. Whether the post-*Furman* reforms worked, the country just changed, or some combination of these or other factors, we can now say that the frequent charge that the death penalty is biased against black defendants is unsupported by the evidence.

What of the claimed “race-of-the-victim bias”? The basis of the claim, as expressed by Justice Brennan, is that “diminished willingness to render [a death] sentence when blacks are victims, reflects a devaluation of the lives of black persons.”⁹⁶ Obviously, a premise of this argument is that imposing the death penalty constitutes valuing the life of the victim. Good, I’m glad we are agreed on that. If we do have a problem, should we fix it by devaluing the lives of more victims (imposing the death penalty less often or never) or by valuing the lives of more victims (imposing the death penalty more often)?

To the extent that a race-of-the-victim disparity exists, is it due to racial animus of the decision makers? It is difficult to see any model of their thought processes consistent with both this hypothesis and the available data. The same prosecutors make the charging decisions in black-victim cases and black-perpetrator cases. How is it possible that a group of people who refrain from seeking the death penalty in black-victim cases because they value black life less fail to seek it more often in black-perpetrator cases for the same reason? If racial animus were behind the claimed disparities, we should see the effect across the board, yet we do not.

The facts that the death penalty is sought and imposed less often in jurisdictions with high black populations and that the poll data indicate much greater opposition to the death penalty among black Americans point to a very different conclusion. The difference in the numbers is not the result of discrimination against black people but rather the result of empowerment of black people. The days of racial exclusion from voting and jury service are long behind us. In localities with a substantial black population, that population has clout in the election of prosecutors and in the verdicts of juries. The exercise of that clout by the only demographic segment of America with a majority opposed to the death

⁹⁶ *McCleskey v. Kemp*, 481 U.S. 279, 336 (1987) (Brennan, J., dissenting).

penalty means fewer death sentences.

Is this a good result? In one sense, it is good to see local democracy and jury of the vicinage working as designed. In terms of the justice of the cases and protection of the community, though, this is a bad result. If an effective death penalty saves lives through deterrence, and there is good reason to believe it does,⁹⁷ insufficient application of it costs a community dearly in unnecessary loss of life. Aside from deterrence, failure to impose the death penalty in those especially heinous cases where any lesser punishment is inadequate costs society in a less tangible but still real way.

Can anything be done? Building public confidence, especially among black Americans, that the death penalty is, in fact, being administered fairly would help. Lack of confidence is doubtless a large part of the reason why opposition is so much higher among black Americans than any other group. Regrettably, the picture the public has been getting is far different from the reality. Those who seek justice for the very worst crimes will have to devote more attention to educating the public on this important topic.

⁹⁷ See Hashem Dezhbakhsh & Paul H. Rubin, *From the 'Econometrics of Capital Punishment' to the 'Capital Punishment' of Econometrics: On the Use and Abuse of Sensitivity Analysis*, 43 APPLIED ECON. 3655 (2011) (answering criticism of their prior paper and showing result is robust); Dale O. Cloninger & Roberto Marchesini, *Reflections on a critique* 16 APPLIED ECON. LETTERS 1709 (2009) (same); Paul R. Zimmerman, *Statistical Variability and the Deterrent Effect of the Death Penalty*, 11 AM. LAW & ECON. REV. 370 (2009) (same).

