SPECIAL ISSUE ARTICLE

WILEY

The use of statistics in criminal cases: An introduction

Christopher Slobogin

Milton Underwood Professor of Law,
Vanderbilt University, Nashville, TN, USA

Correspondence
Christopher Slobogin, Milton Underwood
Professor of Law, Vanderbilt University, 321
21st Ave. S., Nashville, TN 37203, USA.
Email: c.slobogin@vanderbilt.edu

Statistics are increasingly playing a crucial role in criminal cases. As the articles in this Special Issue illustrate, they can form the principal basis for expert testimony (e.g., with respect to risk assessments), heavily influence decisions about policy (e.g., constitutional doctrine governing police stop-and-frisk practices), and help courts evaluate the admissibility of expert testimony (e.g., via proficiency testing). Unfortunately, courts have not always been up to the task of analyzing the value of statistical analysis. In the course of introducing the articles in this issue, this article provides some examples of how statistical information can be misconstrued and suggests several reasons why judges may be reluctant to engage with such information.

1 | INTRODUCTION

The law needs statistics. Statistical analysis helps courts figure out what is known, as opposed to what is merely conjectured. It can confirm or disconfirm otherwise speculative assertions about the important behavioral assumptions the law makes. In many areas of law, without statistics the courts are just guessing.

A scan of the latest edition of the classic casebook by John Monahan and Laurens Walker, *Social Science in Law* (2017), makes clear that courts and legislatures are increasingly realizing that fact. The newest edition is full of legal cases and statutes that rely, at least in part, on statistical information in gauging the effects of pornography, the deterrent impact of the criminal law, the deterrent effect of the exclusionary rule, the dangerousness and culpability of criminal defendants, and a host of other subjects. The implicit assertion of the Monahan and Walker book is that, without such information, legislatures and courts will base their policies and opinions on unsupported folk wisdom, seat-of-the-pants assessments or, worst of all, knowingly biased or prejudiced assumptions about human behavior.

The other side of the coin, demonstrated by the articles in this Special Issue, is that statistical analysis can be misunderstood, misleading, incomplete, or simply wrong (failings that have particularly grave consequences in the criminal setting, the focus of this issue). Even findings that result from sound methodology can be mis-analyzed or misconstrued. Unfortunately, without an understanding of statistics, both lawyers and clinicians can misinterpret or overinterpret research findings.

2 | EXAMPLES OF STATISTICAL MISCONSTRUCTION

One example of this phenomenon in the criminal setting comes from juvenile justice. Social scientists and neuroscientists can justifiably pat themselves on the back for helping to support successful arguments that juveniles are different from adults in legally relevant ways. This research has bolstered Supreme Court opinions that have exempted juvenile offenders from both the death penalty and mandatory life without parole sentences (see, e.g., *Roper v. Simmons* (2005); *Miller v. Alabama*, 2012).

Now, however, the advocacy effort is swinging toward juvenile sentencing generally. Most of the people resentenced after the Court's juvenile-friendly opinions are still receiving extremely long sentences of 25–45 years (Slobogin, 2013, p. 111). Academic advocates argue that these sentences too are unconstitutional, or at least bad policy (see, e.g., Garabedian, 2017, p. 203). In support, they point to research showing that youths under 18 are less psychosocially mature than young adults (see, e.g., Caufmann & Steinberg, 2000). That finding, based on meticulous methodology and boasting good construct validity, confirms intuitions and the Court's decisions to date.

What the research does not do, however, is support the further proposition that juvenile sentences should therefore be drastically shorter than those of adults, the position the research is used to support. Some advocates endorse the stance taken by the American Bar Association's Juvenile Justice Standards, which in commentary state that, given their diminished culpability, juveniles who commit an offense that would result in a 20-year sentence for an adult should receive only a 3-year sentence (ABA, 1980). Others have argued that 16-year-olds are entitled to a 40–50% discount from the adult sentence (Feld, 1997).

From an empirical point of view, these arguments are hard to sustain. While the research indicates that the differences in psychosocial maturity between adults and juveniles are statistically significant, the effect sizes in this research are very small (Slobogin, 2013, p. 112 n. 67). The research does not come close to justifying a 50% reduction, much less an 80% reduction (the ABA's position), in sentences for mid-adolescents. In other words, close statistical analysis of the findings indicates that kids are different, but not *that* different, from adults. As James Andretta and his colleagues emphasize in their article for this issue, statistical significance is not the be-all and end-all of statistical analysis; consideration of effect size is also important (Andretta, Moran, Cantone, & Renbarger, this issue, pp. 133–144). Unfortunately scholars and courts sometimes do not understand that basic point.

A second example of statistical misconstrual involves a much older issue, predictions of violence. Many will remember the famous assertion that, because research showed that psychiatrists and psychologists are more often wrong than right when they opine that a person will reoffend, clinical predictions of dangerousness are worse than predictions based on the flip of a coin (Ennis & Litwack, 1974). This assertion became common enough that it was repeated by Justice Harry Blackmun in his dissent in *Barefoot v. Estelle* (1983, p. 931).

The problem with that assertion, of course, is that it does not take into account base rates. The coin flip analogy would be accurate only if the base rate for violence in the population of interest were 50%. In fact, the base rate for the populations involved in most prediction research is usually much lower, meaning that the predictions made by clinicians, while associated with high false-positive rates, are still much better than chance (Slobogin, 1984, pp. 110–17). Today, this insight is commonplace, and violence predictions are assessed using the receiver operating characteristic and other relativistic analytical techniques. But for a time scholars argued, in some cases successfully, against all judicial use of prediction evidence, based on the statistical misconception that predictions were complete guesswork (see, e.g., *People v. Murtishaw*, 1981, pp. 466–467). This is another example of statistical ignorance on the part of lawyers and courts.

3 | THIS SPECIAL ISSUE

The articles in this Special Issue on the use of statistics in criminal cases provide further examples of why statistical analysis is necessary if the law is to base its assumptions on solid ground rather than on speculation. In "Applying

Statistics to the Gatekeeping of Expert Evidence: Introducing the Structure Statistical Judgement (SSJ)", James Andretta, Grant Moran, Jason Cantone and Rachel Renbarger describe a number of common statistical errors, errors that are well known to sophisticated statisticians but are routinely committed by courts (Andretta et al., this issue, pp. 133–144). Among them are reliance on statistical significance as the sole indicator of relevance, failure to examine effect sizes or conduct power analysis, and misinterpretation of confidence intervals. But judges and legal scholars are not the only ones who fall short in this regard. Andretta et al.'s survey of articles in psychology journals such as *Law and Human Behavior* and *Psychology, Public Policy & Law* reveals a similar lack of statistical rigor in much of the literature written by mental health professionals. The authors' suggestion, which is one echoed in other papers in this issue, is to have court-appointed experts assist judges in evaluating the validity of empirical information presented to the court. They helpfully develop a Structured Statistical Judgement tool to aid in that endeavor (Andretta et al., pp. 133–144).

The other articles in this Special Issue focus on particular uses of statistics in the criminal justice setting. Melissa Hamilton's article, "The Sexist Algorithm", looks at one of the more recent manifestations of statistical analysis – the algorithm (Hamilton, pp. 145–157). Hamilton focuses on the COMPAS, a well-known algorithmic instrument that attempts to assess risk for recidivism. While statistically derived risk assessment instruments can, in theory, reduce bias and subjective decision-making, Hamilton's analysis of the COMPAS shows how such instruments can also result in skewed findings, in this case overclassification of women as high risk. She demonstrates that this effect occurs regardless of how one defines algorithmic fairness.

Nicholas Scurich and Richard John's article, "The Dark Figure of Sexual Recidivism", also concerns risk assessment instruments, but points to a significant reason why those instruments might *underestimate* risk (Scurich & John, this issue, pp. 158–175). Building on the intuition that much crime is not reported or, if reported, not prosecuted, they provide sophisticated statistical proof that current estimates of sexual recidivism seriously undercount the number of crimes committed by sex offenders. This finding has significant implications not only for policymakers constructing legal interventions for sex offenders, but also for researchers who develop risk assessment instruments, at least to the extent that their outcome measure consists solely of arrests or convictions. If statistical analysis is to play a useful and accurate role in assessing risk, the developers of risk assessment instruments will need to be attentive to a wide array of factors, including those identified by Scurich and John.

The Hamilton and Scurich & John articles deal with the role of statistics during the dispositional phase of a criminal case. Statistical analysis can also be useful in evaluating pretrial and trial issues in criminal cases. For instance, as Susan Bandes and her colleagues indicate in their article, "The Mismeasure of Terry Stops: Assessing the Psychological and Emotional Harms of Stop and Frisk to Individuals and Communities", statistics have played a very significant role in how the courts and policymakers view police use of the stop and frisk police tactic (Bandes, Pryor, Kerrison, & Goff, this issue, pp. 176–194). The most prominent statistical analysis in this context has focused on the "hit rate" – that is, the rate at which stops end up producing evidence of criminal activity – and has revealed not only that the hit rate is low overall, but that it is particularly low when people of color are the target. Bandes et al.'s contribution is the insight that statistical analysis is also relevant in evaluating the psychological and emotional harm visited on those who are subject to stops. The authors' suggested outcome measures of this aspect of the stop and frisk practice include its impact on people who have a mental disability or have experienced sexual trauma, its long-term effects on people who are stopped (many of them multiple times), and its effect on the community as a whole. They argue that statistical information about these factors, some of which they provide, should always be considered in determining how to regulate police investigative techniques on the street.

Finally, Gregory Mitchell and Brandon Garrett examine the effect of statistical information on juries (Mitchell & Garrett, this issue, pp. 195–210). Specifically, they look at the influence of proficiency data indicating the skill of an expert at correctly assessing forensic evidence such as fingerprints. The hope of reformers is that apprizing judges or jurors about this information will produce a more nuanced assessment of the expert's accuracy. In fact, Mitchell and Garrett found this assumption to be true only if the jury is told that the expert has a low proficiency rating; when, instead, the jury is informed that the expert has a high proficiency rating, they found the impact on the jury to be no

different than if jurors had not been told anything about proficiency. This finding provides another rationale for using statistics in criminal cases. Without data about proficiency, jurors may be likely to assume that every expert is good at what they do.

At the same time, the authors found that, even when provided error rate information, jurors may discount it "when confronted with an experienced expert with an impressive background, or when other evidence in the case corroborates the expert's conclusions" (Mitchell & Garrett, this issue, pp. 195–210). They also find that people who are higher in objective numeracy gave the fingerprint evidence significantly more weight than those with lower objective numeracy. In short, the value of giving statistical information about expertise to juries remains a subject that can benefit from further study.

4 | WHY STATISTICAL ANALYSIS IS RARE IN LEGAL CASES

The type of analysis one sees in this Special Issue – carried out by lawyers as well as social scientists – rarely finds its way into the caselaw. Admittedly judicial use of social science research has vastly increased and improved over the past quarter-century (Monahan & Walker, 2011). Nonetheless, the main impression one gets from the pages of the Monahan and Walker (2017) casebook is that lawyers and judges are far more likely to ignore the best statistical evidence or, as illustrated by the juvenile justice and risk assessment examples given earlier, misuse or misconstrue the relevant evidence.

Why is that? Why might articles like the ones in this issue end up having little impact in the legal world? One way of answering that question is to look at the views of Supreme Court Justice Lewis F. Powell on the value of statistics in resolving legal cases. Although these views were expressed some time ago, and several other writers have already made note of one or more of them, bringing them together, as they are here, provides additional useful insights into how many lawyers still think about statistics.

One reason why lawyers might ignore statistical information is that they don't trust it. That sentiment was perhaps best expressed at the Supreme Court level by Justice Powell in the case of *Ballew v. Georgia* (1978), which involved consideration of whether a five-person jury is constitutional. Justice Blackmun, for the majority, relied on social science evidence indicating that juries of less than six members tend to deliberate less effectively and to ignore minority viewpoints. Justice Powell, in dissent, would have none of it. He wrote, "I have reservations as to the wisdom—as well as the necessity—of Mr. Justice Blackmun's heavy reliance on numerology derived from statistical studies" (p. 246). Numerology is the study – common in primitive societies – of numbers; for instance, one might examine the figures in a birth date to glean their influence on human affairs. Read literally, then, in *Ballew*, Justice Powell was saying that the scientific research on which Justice Blackmun relied was mumbo-jumbo. Few lawyers today are going to be that direct. But Justice Powell's candid comment probably accurately reflects the hostility toward statistical information that some members of the bar still harbor.

A second reason why lawyers are leery of statistics, closely related to the first, is that they don't understand them. Whatever the truth of the old saw that lawyers are lawyers because they did not like math and the sciences, it is clear that most lawyers have not had special training in statistics. Again Justice Powell provides an example of this phenomenon, this time in connection with his majority opinion in *McKleskey v. Kemp* (1987), which rejected the argument that the death penalty is implemented in a racially discriminatory fashion. In the course of justifying this holding, Powell's majority opinion simply refused to apply the normal rules of statistical inference. Instead, the justice summarily dismissed the findings of the extremely comprehensive Baldus, Pulaski, & Woodworth (1983) study proffered by the defendant, a study that concluded that the race of the victim is a primary determinant of who receives the death penalty. As Powell's biographer later reported, one reason he did so was that "[m]y understanding of statistical analysis ... ranges from limited to zero" (Jeffries, 1994, p. 49).

These first two reasons as to lawyers find statistics unpalatable – the beliefs that they are mere numerical hocuspocus or completely incomprehensible – cannot be given any credence. The third reason, by contrast, has some

strength to it. It is the notion that empirical information is of doubtful relevance to the legal system because it is nomothetic (i.e., based on group data), whereas the job of lawsuits and the courts is to decide individual cases. In work I have done with Monahan and David Faigman, we have labeled this the G2i (general-to-individual) problem (Faigman, Monahan, & Slobogin, 2014). As the late Erica Beecher-Monas stated, "Statistical inference from the general to the specific is simply not something science can do" (Beecher-Monas, 2014, p. 1077).

Justice Powell once again provides an illustration of the point. In a memo addressed to his fellow justices, a second reason Powell gave for dismissing the Baldus et al. study in *McCleskey* was that "sentencing judges and juries are constitutionally *required* to consider a host of individual-specific circumstances in deciding to impose capital punishment. No study can take all of these individual circumstances into account, precisely because they are fact-specific as to each defendant" (Sundby, 2012, p. 12, emphasis in original). He is right, but as the articles in this Special Issue attest, that does not mean that statistical analysis based on groups is irrelevant to legal cases. If generalized information could not be used to resolve disputes about individuals, no expert could testify about anything (Faigman et al., 2014).

A fourth reason why courts resist statistics is practical rather than conceptual. Judges fear they will get bogged down in mind-numbing debates about statistical significance, effect sizes, and power analysis, when, in their view, anecdotal evidence can provide the necessary information. Echoes of that sentiment are found in Justice Powell's dissent in *Castanada v. Partida* (1977), where the majority relied almost entirely on statistical evidence to support its holding that a Texas county's grand jury selection process discriminated against Hispanics. Pointing out that many of the county's officials were Hispanic, Powell said, "There is for me a sense of unreality when Justices here in Washington decide solely on the basis of inferences from statistics that the Mexican-Americans who control the levers of power in this remote border county are manipulating them to discriminate against themselves" (p. 518). In other words, Justice Powell felt that the Court should rely on common sense, not standard deviations, to figure out when discrimination has occurred.

As this last quote from Powell also suggests, a final reason lawyers do not want to focus on statistics is because doing so would occasion a substantial relinquishment of their power. Until scientists came along, legislators and judges controlled decisions about what is real and what is not, how the reasonable person acts, what the average person thinks, and how much harm a particular act causes. To the extent that lawyers have to consult scientists about these types of issues, they are surrendering that prerogative (see, generally, Faigman, 2008, p. 16). Of course, rather than resisting that development, courts could welcome help with the hard decisions they have to make. Their reluctance to do so should be assuaged by the fact that, once the facts are established, they, not the scientists, are entitled to – and in fact must – make the normative judgments that legal cases almost always require.

5 | CONCLUSION

Lawyers might share a number of phobias about statistical analysis, ranging from the beliefs that it is quackery or incomprehensible to the views that it is irrelevant, unnecessary, or a subterfuge aimed at taking over the system. And lawyers may not be the only professionals with these types of reaction. It is probably fair to say that many mental health professionals have them as well. Many clinicians would probably rather rely on their seat-of-the-pants assessment than on numbers, because they are not sure they understand statistics, think the numbers might be confusing to the factfinder, believe they know more about the individual case in front of them than statistics could ever reveal, or simply would prefer that the arbiter of the issue be them, not the numbers.

Whether held by lawyers or clinicians, the first two of the five concerns I've identified – the belief that statistics are merely numerology and the fear of grappling with statistical analysis – can be rectified by articles of the type that appear in this Special Issue. And if those concerns are alleviated, perhaps the last three obstacles – the beliefs that, with respect to legal matters, statistical information is immaterial, surplusage, or a surrender of power – will be pushed aside as well.

REFERENCES

American Bar Association (1980). *Standards Relating to Juvenile Delinquency and Sanctions*. Chicago: American Bar Association.

Andretta, J. R., Moran, G., Cantone, J. A., & Renbarger, R. (2019). Applying statistics to the gatekeeping of expert evidence: Introducing the structured statistical judgement (SSJ). *Behavioral Sciences & the Law*, *37*, 1–12.

Baldus, D. C., Pulaski, C. A., & Woodworth, G. G. (1983). Comparative review of death sentences: An empirical study of the Georgia experience. *The Journal of Criminal Law and Criminology*, *74*, 661–753. https://doi.org/10.2307/1143133

Ballew v. Georgia (1978). 435 U.S. 223.

Bandes, S. A., Pryor, M., Kerrison, E. M., & Goff, P. A. (2019). The mismeasure of *Terry* stops: Assessing the psychological and emotional harms of stop and frisk to individuals and communities. *Behavioral Sciences & the Law*, *37*, 1–19.

Barefoot v. Estelle (1983). 463 U.S. 880.

Beecher-Monas, E. (2014). Lost in translation: Statistical inference in court. *Arizona State Law Journal*, *46*, 1057–1102.

Castaneda v. Partida (1977). 430 U.S. 482.

Caufmann, E., & Steinberg, L. (2000). (Im)maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences & the Law*, *18*, 741–760. https://doi.org/10.1002/bsl.416

Ennis, B. J., & Litwack, T. R. (1974). Psychiatry and the presumption of expertise: Flipping coins in the courtroom. *California Law Review*, *62*, 693–752. https://doi.org/10.2307/3479746

Faigman, D. L. (2008). *Constitutional Fictions: A Unified Theory of Constitutional Facts*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341270.001.0001

Faigman, D. L., Monahan, J. T., & Slobogin, C. (2014). Group to individual (G2i) inference in scientific expert testimony. *University Of Chicago Law Review*, *81*, 417–480.

Feld, B. C. (1997). Abolish the juvenile court: Youthfulness, criminal responsibility, and sentencing policy. *The Journal of Criminal Law and Criminology*, *88*, 68–136. https://doi.org/10.2307/1144075

Garabedian, C. (2017). Juvenile empiricism: Approaches to juvenile sentencing in light of Graham and Miller. *U.C. Davis Journal of Juvenile Law & Policy*, *21*, 195–214.

Hamilton, M. (2019). The sexist algorithm. *Behavioral Sciences & the Law*, *37*, 1–13.

Jeffries, J. C. (1994). *Justice Lewis F. Powell, Jr.: A Biography*. New York, N.Y.: Fordham University Press.

McCleskey v. Kemp (1987). 481 U.S. 279.

Miller v. Alabama (2012). 567 U.S. 460.

Mitchell, G., & Garrett, B. (2019). The impact of proficiency testing information and error aversions on the weigh given to fingerprint evidence. *Behavioral Sciences & the Law*, *37*, 1–16.

Monahan, J. T., & Walker, L. (2011). Twenty-five years of social science. *Law and Human Behavior*, *35*, 72–82. https://doi.org/10.1007/s10979-009-9214-8

Monahan, J. T., & Walker, L. (2017). *Social Science in Law* (9th ed.). New York: N.Y.: Thomson Reuters.

People v. Murtishaw (1981). 631 P.2d 446 (Cal.).

Roper v. Simmons (2005). 543 U.S. 551.

Scurich, N., & John, R. S. (2019). The dark figure of sexual recidivism. *Behavioral Sciences & the Law*, *37*, 1–18.

Slobogin, C. (1984). Dangerousness and expertise. *University of Pennsylvania Law Review*, *133*, 97–173. https://doi.org/10.2307/3311865

Slobogin, C. (2013). Treating juveniles like juveniles: Getting rid of transfer and expanded adult jurisdiction. *Texas Tech Law Review*, *46*, 103–141.

Sundby, S. (2012). The loss of constitutional faith: *McCleskey v. Kemp* and the dark side of procedure. *Ohio State Journal of Criminal Law*, *10*, 5–33.