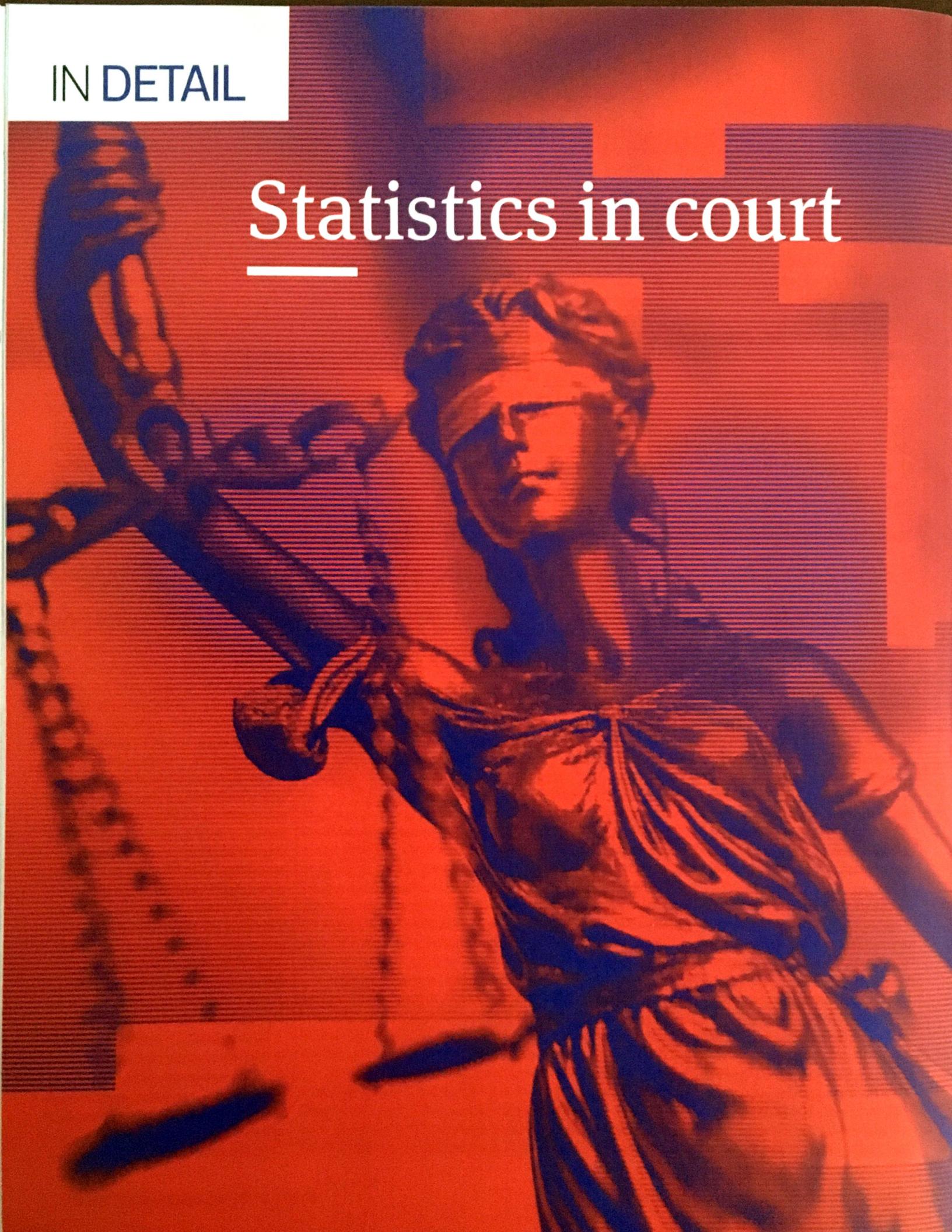


IN DETAIL

Statistics in court



Judges and jurors are often asked to make sense of statistics. But data, probabilities and uncertainties are easily misunderstood or misused by those not trained to deal with them. Is education the answer? Or is greater oversight required? **Nick Thieme** considers the options

An elderly woman is picking up litter when she is barrelled over and robbed by a younger white woman with a blond ponytail. A few moments later, his attention drawn by crying and screaming, an eyewitness sees a woman matching that description hop into a yellow car driven by a black man with a beard and a moustache. Four days later, Janet Collins, a white woman with a blond ponytail, and her husband Malcolm Collins, a black man without a beard, are questioned about the crime. Two weeks later they are charged with the robbery and four months later they are convicted.

A key piece of evidence in their trial was an "analysis" performed by a mathematics instructor at a state college. It went like this: the probability of a partly yellow automobile is $1/10$, of a man with a moustache is $1/4$, of a woman with a ponytail is $1/10$, of a woman with blonde hair is $1/3$, of a black man with a beard is $1/10$, and of an interracial couple in a car is $1/1000$. Multiplying these together, the prosecutor concluded that "there was but one chance in 12 million that any couple possessed the distinctive characteristics of the defendants", and that "the chances of anyone besides these defendants being there, ... having every similarity, ... is something like one in a billion".

The absolute intellectual bankruptcy of this argument is egregious. According to *The Double Helix and the Law of Evidence*, by David H. Kaye, "the court saw two 'glaring defects' in the calculation – 'an inadequate evidentiary foundation' for the numbers that were multiplied and 'an inadequate proof of statistical independence'" in the physical traits in question, pointing out that black men with beards and men with moustaches represent "overlapping categories".¹ The California Supreme Court overturned the case on appeal. But mistakes of this sort are still made in courtrooms today. Why should that be, what can be done to limit them, and how do we guarantee defendants a fair shake when statistics are used in court?

The limits of education

It would be uncharitable and incorrect to imply that statistics are always misused or misunderstood in a legal setting. Indeed, the US legal system's statistical highlights are as prescient as they are exemplary. One of the best writers to ever serve on the Supreme Court, Associate Justice (and Acting Chief Justice) Oliver Wendell Holmes, Jr, wrote in 1897 that: "For the rational study of the law the black letter man may be the man of the present, but the man of the future is the man of statistics and the master of economics." This would be proven right time and time again.



Nick Thieme is a freelance reporter, with writings appearing in *Significance*, *Slate*, *BuzzFeed*, *NOVA*, and *Undark*. He is also a researcher with New America's Open Technology Institute, investigating fair internet usage and net neutrality.

In a case addressing the voting rights of black Alabamians under the Civil Rights Act of 1957, the Supreme Court was informed by statistical analysis and voted in favour of black citizens, saying: "In the problem of racial discrimination, statistics often tell much, and Courts listen." In 1986, *Bazemore v Friday* established that regression analysis can be used in some cases to determine racial discrimination in the workplace. And in 2011, *Matrixx v Siracusano* set out that, roughly speaking, statistical significance is not a requirement for practical significance.² Explaining the court's opinion, Associate Justice Sonya Sotomayor stated that "courts frequently permit expert testimony on causation based on evidence other than statistical significance". This understanding came to the court years before it dawned on parts of the scientific community.

However, while judges often get statistics, they and others do not always get it right. Jury trials, a hallmark of the American legal system, provide difficulties of their own. We might expect judges or lawyers to be familiar with basic statistics, or areas of statistics that are particularly relevant to their profession, such as forensic statistics. But forensic statistics can present challenges even to those well trained in statistics, says Paul Weiss, a faculty researcher at Emory University, who testifies as an expert witness. So, he asks: "How would you expect 12 people from any walk of life to get it?"

Education can help, but requiring that judges and jurors be statistically literate is no panacea for the misuse of statistics in the law. If only the statistically literate can be jurors, there would be no way to guarantee a defendant their 6th Amendment right to a jury of their peers. So that idea is a non-starter, barring a wonderful, watershed change to universalise statistics education.

Constitutional issue aside, it should be noted that people with statistical education can and do use statistics incorrectly, as evidenced by the replication crisis looming over the sciences. The Royal Statistical Society's publication, *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings*, puts an even finer point on the issue: "twisted logic can seem enormously seductive and [is] frequently perpetrated by professionals who ought to know better, especially in pressured situations such as giving evidence in criminal trials" (bit.ly/2Pp3ZbY).

Errors and fallacies

Statistical errors often arise in court because a statistic is proffered as an answer to a question it cannot address. Consider how juries might interpret the following phrases: "evidence is consistent with", "could have come from", or "does not rule out". When an expert testifies that "the evidence is consistent with the defendant having murdered the victim", they mean the data does not rule out the defendant having committed the murder. Jurors and judges may sometimes take that to mean the defendant likely murdered the victim.

The well-known prosecutor's fallacy is another example of using a statistic to answer a question it cannot. Many famous cases have been decided by experts confusing the probability ►

► of damning evidence given a defendant's innocence with the probability of innocence given damning evidence.³ Expert witnesses may testify to the first – "If the suspect is innocent, it would be unlikely to find gun residue on their hands" – but do so meaning the second – "Therefore, the probability this person is innocent given the evidence is very low".

Dr Maria Cuellar, assistant professor of criminology at the University of Pennsylvania, has studied and testified on shaken baby syndrome, and she described the problem with testifying in her field like this: "[We're] using statistical arguments within medicine to talk about legal questions. That makes it very confusing because the standards required are different in the different fields, and there's very little communication between them."

That the standards are different can again be framed as statistics being used to answer the wrong question. Consider an example Cuellar suggested. A worker sues a chemical company after 10 years of employment because they contracted lung cancer just before retiring. A statistician determining the carcinogenic effects of the chemical produced by the company and inhaled by the worker is searching for the probability that exposure to the chemical causes cancer, and may testify on that basis if called as an expert witness once the case goes to trial. However, in deciding the case, the judge is interested in a subtly different question: whether the worker's exposure to the chemical in this factory caused this particular cancer. The statistician's question is one of causal effect, while the judge's is one of causal attribution.⁴

There are, of course, statistical methods for this kind of causal reasoning. Bayesian methods, including causal belief networks, causal inference specifically for reasoning about legal issues, and randomised controlled trials are all active fields of study. If the issue were purely the unavailability of models, there would be no issue. The issue must be elsewhere.

Conflicting notions

One reason why it can be so easy to demand the wrong legal answers of the wrong statistics is because the fields, as practised, represent competing views on how to reason about the truth. Dr Jim Greiner, a professor of statistics and law at Harvard, with experience as a trial attorney, explained: "There's an epistemology within the United States legal profession that says what's true is what I see day to day and what [higher courts] tell me is true." Statistical reasoning, on the other hand, largely discounts individual cases and assigns less value to hierarchy. Each philosophy was developed with a goal in mind and each has benefits, but their differences can cause problems in integration.

The clearest example of this is their attitudes towards precedent. In law, the longer a rule has been adhered to the more correct it is considered to be, hence why the US Constitution is viewed as the ultimate tool for reasoning about the law of the land. On the other hand, science – and statistics as a part of it – is an iterative process that improves with time. The newest methods are typically those that allow us to reason most effectively. These two different notions conflict in practice.

The curious case of Mats Järström

Mats Järström is an engineer. He repaired aircraft for the Swedish Air Force and designed speaker systems for American customers. But for the past four years he has been in and out of courtrooms because of his work on traffic light signals in Beaverton, Oregon. Järström's curiosity about the lights began in May 2013 when his wife received a citation for running a red by 0.12 seconds. Having spent decades working with margins of error in electronic timing, this hairline violation caught his eye. Together, the Järströms videotaped hours of footage intended to measure the length of the yellow light timings in Beaverton. They discovered that almost all of them were shorter than the amount of time required by law. But the state of Oregon was not impressed by the research. Read the full story at bit.ly/jarlstrom.

In the 1986 case of *Thornburg v Gingles*, the Supreme Court accepted the results of two statistical models in determining "racially polarised voting" as a form of vote dilution. Because of the court's influence, those two models have remained the standard for detecting racially polarised voting in the 30 years since, despite the existence of far better models. Courts' reliance on these methods is so all-consuming that a research project headed by Greiner was unable to discover any instance of a court finding racial bloc voting when the plaintiff did not present evidence using one of the two suggested models.⁵

Improved statistical methods strive to make data analysis easier and inference more thorough, but they are unlikely to singlehandedly bridge the divide between law and statistical science. The more advanced statistical methods become, the more they may represent an unfamiliar way of reasoning, and the more difficult they may be for judges, lawyers and jurors to understand.

Perhaps the solution is to recognise that competing methods for reasoning can be made complementary and that a balance can be found between satisfying statistical and legal concerns. Where this balance is achieved, statisticians could help experts and attorneys to give explanations better suited to the aims of the law.

On the part of the statistician, the courts require methods that are interpretable and interrogable by lay people, as well as honesty about the limitations of those methods. On the legal side, the courts need lawyers who are willing to listen to and learn from their experts.

This is also where statistical education could have a greater effect on legal outcomes. Lawyers better trained in statistics and statistical reasoning will be better prepared to take advantage of the work done by their mathematical allies. But if attorneys shop around for "statistical experts" until they find someone willing to tweak sample sizes and sampling procedures to get the desired result, no reconciliation is possible.

Expert oversight

What if there were to be an Inspector General-like body tasked with overseeing statistical arguments in legal settings? Such an authority would review the use of survey sampling, hypothesis testing, and model selection in courts, recommending, if necessary, that judges dismiss evidence or litigants appeal. Statistical oversight of this sort could help judges in their role as scientific gatekeepers – especially now that DNA tests and forensic science are commonly presented as evidence.

Almost all the experts interviewed for this article emphasised the difficulty of spreading scientific knowledge into the legal system. Greiner says the diffusion of information is "a problem about governing in a world of intensely specialized knowledge" – but this is a role for which an oversight body is particularly well suited. Consider an analogy to medicine. Not all doctors are intimately familiar with the chemical structure of the medicines they prescribe because, frankly, chemistry is not their primary concern. For that reason, doctors rely on the chemists at the Food and Drug Administration to classify drugs

Statistical errors often arise in court because a statistic is proffered as an answer to a question it cannot address

and drug interactions as safe or not. A hypothetical statistical legal agency could do the same.

However, courts may be reluctant to cede this considerable power to an outside body. As put by Supreme Court Associate Justice Stephen Breyer (bit.ly/2N6MHv9): "Any effort to bring better science into the courtroom must respect the jury's constitutionally specified role – even if doing so means that, from a scientific perspective, an incorrect result is sometimes produced." It is unsettling, but Breyer's argument – in essence – says that juries must be allowed to make mistakes in reasoning about statistics as they do in reasoning about medicine, physics, and any other science presented to them. Justice is not only for the experts to decide.

Political considerations

The freedom of juries need not come at the expense of better science and statistics in court. In many countries, including the UK, independent regulators set standards for forensic science, and those standards demand that methods and results be scientifically and statistically valid. These regulators also establish the boundaries of what expert witnesses can and cannot say about a piece of evidence.

In the US there is no national oversight of forensic labs and techniques. Instead, the Center for Statistics and Applications in Forensic Science (CSAFE) aims to partly fill that gap by developing and refining forensic techniques and providing educational opportunities for practitioners. But CSAFE's director, Dr Alicia Carriquiry – distinguished professor of statistics at Iowa State University – worries about the inconsistency with which US administrations commit to addressing the misuse of statistics and science in court. She says that: "In the Obama administration, there was a great deal of cooperation between the Department of Justice and the research community, and a

whole lot of support for basic sciences geared toward improving the scientific and statistical foundations of forensic tools. But the Trump administration came along and many of these efforts came to a screeching halt."

CSAFE's work and funding, along with the work and funding of other groups that conduct research into the accuracy of statistical methods used in court, has been hampered and in some cases stopped under the new administration.

This political ebb and flow of effort and funds does not seem to happen to such an extent with organisations like the National Science Foundation and the National Institutes of Health. And yet, as Carriquiry puts it: "The principled use of statistics is crucial for the fair administration of justice."

If fair justice is something we as a society are interested in, then putting a stop to the misuse and misunderstanding of statistics in court will require more consistent attention from all branches of government. In the absence of that, Carriquiry encourages statisticians to "find your way to the closest Innocence Project office or public defender's office and offer your services as a statistical consultant". ■

References

1. Kaye, D. H. (2010) *The Double Helix and the Law of Evidence*. Cambridge, MA: Harvard University Press.
2. Ziliak, S. T. (2011) *Matrixx v. Siracusano and Student v. Fisher*. *Significance*, **8**, 131–134.
3. Skorupski, W. P. and Wainer, H. (2015) The Bayesian flip: Correcting the prosecutor's fallacy. *Significance*, **12**(4), 16–20.
4. Yamamoto, T. (2012) Understanding the past: Statistical analysis of causal attribution. *American Journal of Political Science* **56**(1), 237–256.
5. Greiner, D. J. and Quinn, K. M. (2010) Exit polling and racial bloc voting: Combining individual-level and RxC ecological data. *Annals of Applied Statistics*, **4**(4), 1774–1796.