2000

Statistics, Not Experts

Cass R. Sunstein
William Meadow

Follow this and additional works at: http://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation

This Working Paper is brought to you for free and open access by the Coase-Sandor Institute for Law and Economics at Chicago Unbound. It has been accepted for inclusion in Coase-Sandor Working Paper Series in Law and Economics by an authorized administrator of Chicago Unbound. For more information, please contact unbound@lawuchicago.edu.
STATISTICS, NOT EXPERTS

William Meadow and Cass R. Sunstein

This paper can be downloaded without charge at:

The Chicago Working Paper Series Index:
http://www.law.uchicago.edu/Publications/Working/index.html

The Social Science Research Network Electronic Paper Collection:
Statistics, Not Experts

William Meadow* and Cass R. Sunstein**

Abstract

The legal system should rely much more than it now does on statistical evidence. It should be cautious about the judgments of experts, who make predictable cognitive errors. Like everyone else, experts have a tendency to blunder about risk, a point that has been shown to hold for doctors, whose predictions significantly err in the direction of optimism. We present new evidence that individual doctors’ judgments about the ordinary standard of care are incorrect and excessively optimistic. We also show how this evidence bears on legal determinations of negligence, by doctors and others.

I. Introduction

Trials frequently raise questions that call for expert intervention. Experts are asked to answer a range of questions about what people, or professionals, ordinarily do. If a doctor is accused of negligence, for example, it is necessary to know about the customary practice of doctors. Of course negligence judgments depend, at least in part, on an assessment of ordinary practice.¹ But how is ordinary practice assessed?

The basic answer is that the assessment comes via statements from expert witnesses, describing the ordinary practice.² There can be no doubt that experts know a great deal about topics on which ordinary people lack information. But experts, no less than other people, are subject to predictable biases.³ Their judgments about probability are affected by the same heuristics and biases to which most people are subject, even if (and this is a disputed question) expertise tends to reduce the most serious errors. Our basic proposal here is that the legal system should rely, much more than it now does, on statistical data about doctors’ performance, rather than the opinions of experts about doctors’ performance. For the first time, it is becoming possible for law to rely on this evidence, precisely because such evidence is becoming

* Associate Professor, Pediatrics; Associate Director, Neonatology, Assistant Director, MacLean Center for Clinical Medical Ethics, The University of Chicago
** Karl N. Llewellyn Distinguished Service Professor of Jurisprudence, Law School and Department of Political Science, University of Chicago.
¹ There is a dispute about the extent to which ordinary practice is determinative. For the classic case, see The T.J. Hooper, 60 F.2d 737 (2d Cir. 1932).
² We build on existing law here, without intending to enter into debates about the extent to which law should simply incorporate, or instead sometimes improve on, the existing standard of care. On any view, the customary practice is relevant, and that is sufficient for our purposes here.
³ See Jonathan Baron, Thinking and Deciding (2d ed 1994).
increasingly available. If our argument is convincing in the medical context, it should apply in many other settings in which experts are asked to testify about negligence or deviations from ordinary practices. In many settings, the fallible opinions of isolated experts should be supplemented or replaced with statistical data. A step of this sort would dramatically increase the sense and rationality of tort law.

Our proposal rests partly on the claim that the use of statistical data will greatly simplify litigation and reduce the role of strategic behavior in the litigation process. Battles of isolated experts should be easier to mediate when there is a large pool of evidence on which to draw. But our larger claim is that by using statistical data, the legal system will reach much more accurate results. If the law is seeking the standard of care, it should not depend on fallible memories and recollections of local experiences; it should draw instead on more global evidence of the kind that is now increasingly available for use in court.

II. Excessive Optimism About Risks: In General

It is now well-known that most normal people tend to be risk optimists, in the sense that they believe themselves to be relatively immune from risks that are faced by similarly situated others. This is one of the most robust findings in social psychology. For example, 90% of drivers believe that they are safer than most drivers and less likely to be involved in a serious accident. Most people believe that they are distinctly unlikely to be subject to various risks, such as cancer, heart disease, and divorce. Smokers appear to know the statistical risk of smoking; but they believe that they are less likely than most smokers to fall victim to the various risks. In one study, less than half of smokers believed that they have a higher-than-average risk of cancer or cardiovascular disease; indeed, most heavy smokers (over forty cigarettes per day) believe that they are not at any increased risk.

Only one group of people does not show a tendency to excessive optimism: the clinically depressed. It would be natural to infer that though unrealistic optimism makes for erroneous predictions about outcomes, it is on balance adaptive, especially insofar as an optimistic attitude increases the probability of a good outcome and otherwise creates hedonic benefits.

---

4 Some of the data here is discussed, from a different angle, in William Meadow et al., Ought 'standard care' be the 'standard of care'? A study of the time to administration of antibiotics in children with meningitis, 147 Am J Dis Child 40 (1993).
7 See Weinstein, supra note.
9 John Ayanian and Paul Cleary, Perceived Risks of Cancer and Heart Disease Among Cigarette Smokers, 281 JAMA 1019 (1999).
10 See Taylor, supra note.
11 This is the thesis of id.; see also the discussion of the self-fulfilling prophecy in Nicholas Christakis, Death Foretold: Prophecy and Prognosis in Medical Care 135-62 (2000).
Alternatively, it might be thought that some social settings would work against optimistic bias. In markets, for example, entrepreneurs might have an incentive toward realism, especially because market pressures, it might be expected, would punish, and drive out, those who are unrealistically optimistic. Perhaps markets do move people toward realism; but entrepreneurs, no less than anyone else, have been shown to suffer from unrealistic optimism.\(^{12}\) One might intuit that those with specialized knowledge are less prone to this effect. This appears not to be the case. It has almost universally been found that physicians, no less than others, have a substantial tendency to err in predicting outcomes. Most often the errors are in the direction of optimism.

"Virtually all" of the existing studies of physicians "have documented frequent and large errors in predictions."\(^{13}\) No study finds a high level of accuracy. And the errors tend in a particular direction: "physicians are prone to an optimistic bias."\(^{14}\) As an example, a study in 1972 showed that in making predictions about length of survival for cancer patients, only 47% of the physicians provided prognoses that were even roughly accurate, and 80-90% of the mistakes were excessively optimistic.\(^{15}\) A subsequent study, conducted in 1987, found overestimates of survival time in 88% of the cases, by roughly a factor of three.\(^{16}\) A recent, large-scale study found inaccurate predictions in 80% of cases, with 63% of these showing overestimates.\(^{17}\) Interestingly, patient characteristics (age, race, sex, illness duration) were not correlated with inaccurate, optimistic estimates, but one factor was: "the better the doctor knew the patient - as measured, for example, by the length and intensity of their contact - the more likely the doctor was to err in the prognosis, most frequently by overestimating survival."\(^{18}\) The overall picture is that in cases involving cancer patients, physicians accurately predict survival only 10% to 30% of the time, and the rest of the time they overestimate survival by a factor of two to five.\(^{19}\)

All of this is highly suggestive, but not decisive, with respect to the particular point that we mean to investigate here: whether experts make erroneous judgments about the ordinary standard of medical care, and whether the errors go in a predictable direction. It would be plausible to think that the general tendency toward optimism would affect those judgments as well; but it would also be possible to imagine that (a) doctors’ predictions about their patients’ prospects are systematically overoptimistic, but (b) doctors have an accurate sense of what is ordinarily done by themselves and other doctors. If there is no systematic error with respect to (b), there would be little need to substitute data for experts. We now offer the first real evidence on that question.


\(^{13}\) See Christakis, supra note, at 66.

\(^{14}\) Id.

\(^{15}\) Id. at 66.

\(^{16}\) Id. at 66-67.

\(^{17}\) Id. at 67.

\(^{18}\) Id. at 67-68.

\(^{19}\) Id. at 68.
III. Standard of Care: Experts vs. Data

We now summarize some of our own evidence that doctors err with respect to ordinary conduct in their own fields. As expected, the apparent basis for the magnitude and direction of these errors is optimistic bias.

A. Bacterial Meningitis

Consider the problem of bacterial meningitis in children. For the purposes of the current discussion, it is necessary to know only three things about that problem. Bacterial meningitis is an infection of the brain; it can (usually) be treated with antibiotics; as a general rule, the sooner the antibiotics are started, the better the outcome will be.

In medical malpractice cases, the question often arises whether treatment of a particular child was unduly "delayed." Of course a key issue is how a legal decision-maker would know what kind of delay counts as "undue." For obvious reasons, juries are assumed to be ignorant of the medical facts in such cases, and consequently are instructed to rely on the testimony of expert witnesses. The instruction that jurors should rely on expert opinions for their knowledge about the standard of medical care answers one question, but it raises another -- on what should the experts themselves rely?

In the traditional formulation, experts themselves are instructed to rely on their personal "knowledge and training." But for reasons given above, such admonitions raise many problems. In the context at hand, it would be reasonable to predict that experts would tend to be excessively optimistic – that their recollections of the time that antibiotics are initiated in children with meningitis would be tilted toward shorter values than actually occurred. And if so, experts, relying on their "knowledge and training," will provide inaccurate accounts of behavior.

We tested this hypothesis directly, by surveying two groups of pediatric experts about their estimates of timing of antibiotic administration in children who present with bacterial meningitis (ABTIME), and subsequently comparing these estimates to actual data obtained from real cases of meningitis. We began by identifying 54 experts in pediatric emergency medicine who were attending a national meeting of their subspecialty (most children with bacterial meningitis present to the emergency room, where they receive their initial medical care). We asked each of these experts to give us their estimate of the average time that would elapse from arrival of a child with meningitis in their emergency room to the start of antibiotic therapy for that child. The average estimate of these experts was 56 minutes -- 95% of these experts' opinions fell within the range between 20 - 120 minutes.

We then posed precisely the same question to another group, consisting of 23 experts in the field of pediatric infectious diseases (the subspecialty with most experience in the care and treatment of meningitis for children after admission to the hospital). The average estimate of ABTIME offered by pediatric infectious disease experts was 87 minutes, slightly more than 50% longer than the intuitions of the experts in emergency medicine.
Moreover, and most importantly, the opinions of both of these groups of potential expert
witnesses were **wrong**, and wrong by a large margin. We compared both of these distributions of
expert opinion to actual times, determined on the basis of a review of 93 cases of children with
meningitis seen at two university medical centers in Chicago. The average value of ABTIME for
these 93 children was 120 minutes, with 95% of cases falling in the interval between 30 to 240
minutes. When the medical literature was reviewed in an attempt to extend these observations of
actual ABTIME, 200 reported cases of children with meningitis were found from hospitals in
South Carolina and California, with an average time comparable to those observed at our
Chicago hospitals (median 114-126 minutes). The following table captures the differences:

| Variations between Expert Opinions and Data-based Observations for ABTIME |
|---------------------------|-----------------|----------------|-----------------|
| ID Experts                | ER Experts      | Chicago Hospitals | SC and CA Hospitals |
| 56 minutes                | 87 minutes      | 120 Minutes       | 120 minutes     |

Whatever "expert" opinion means in this context, it does not mean an accurate opinion.
Both groups of potential pediatric experts were simply wrong in their opinions about the time to
begin antibiotics in children with meningitis. The 77 expert estimates of ABTIME were
significantly shorter than the 293 actual determinations of ABTIME to a probability of less than
1 part in 1000. In addition, the inaccuracy noted in the estimates was consistent with our
hypothesis: responses were biased toward the outcome perceived to be 'desired', that is, shorter
waiting periods.

Others, in varying contexts, have demonstrated a similar disconnect between physicians'
description of their own practices and "objective" determinations of the same behavior. As
examples, consider patient education by physicians in general practice. Physicians are
committed to the value of patient education,20 and spend up to 25% of their office time
counseling patients.21 Clearly, the desired practice is to spend this time explicitly informing
patients about health-related behaviors, such as the uncontested virtues of smoking cessation and
preventive oncology. Nevertheless, when physicians' reports of their office practices are
compared with taped interviews of these same encounters, there is only a weak, insignificant
correlation between the self-reported patient education activities and the actual performance.22
Not surprisingly, physician's recollection err in the predicted direction; that is more explicit
counseling is recalled than actually occurred.

20 H Wechsler et al., The physician's role in health promotion -- a survey of primary-care practitioners., 308 N Engl
J Med. 97 (1983); T.E. Kottke et al., Attributes of successful smoking cessation interventions in medical practice. A
21 C.T. Orleans et al., Health promotion in primary care: a survey of US family practitioners, 14 Prev Med 636
(1985).
22 See id; see also J.E. Davis et al, Cancer prevention and screening activities in primary care practice, 16 Prev
Once again, our fears are confirmed --- we expect anecdotal recall to be inaccurate, biased, and inevitably, systematically imperfect, and we find it to be so. If, as the psychological evidence suggests, our findings here are the rule and not the exception, it is hardly clear that the law should continue to base a system of medical-legal jurisprudence on such a shaky foundation.

B. The Proposal

Of course the evidence that we have offered here is suggestive rather than exhaustive. Perhaps there are contexts in which doctors, or other experts, have an accurate sense of the ordinary standard of care. But we believe that general evidence of error, together with the particular evidence introduced here, is sufficient to show that both error and optimism are highly likely to infect a wide range of expert testimony. This point operates independently of the ordinary incentives, in an adversary system, to assist one’s own side, whatever it may be. It therefore makes sense to move toward greater reliance on data, and less reliance on the recollections of isolated experts, certainly in the context of malpractice suits, and probably more generally. Until recently, the legal system has been unable to rely on statistical data for the simple reason that it has not existed. But it is increasingly common to develop data sets about physician choices and behavior, and the legal system will have an increasingly large amount of information on which to draw.

Our proposal would have two large advantages. First, it would simply the task of discerning the truth. It is now exceedingly difficult for juries to decide whether one or another expert has adequately captured general experience. This issue may well turn on unreliable judgments about credibility or on sympathy, not relevant to the standard of care issue, for one or another side. Second, and more important, use of statistical evidence would inevitably increase accuracy, reflecting the comparative advantage of pooled data over individual recollections. There is no good reason for courts to continue to rely on the latter when they have access to the former.

III. Problems and Counter-arguments

A. Statistical Error?

Our proposal might be criticized on the ground that statistical analysis is itself subject to error. Of course there can be no assurance that any particular account of relevant data is correct. If this is so, it might be objected, the use of statistical evidence will merely be the beginning of a continuing battle among imperfectly reliable experts. How much would be gained by that?

This objection has a degree of truth, but rather than impeaching our proposal, it suggests the need to develop good methods for evaluating any particular claim about what the data establishes. There is no escaping tests of the data in the ordinary way, through the presentation of conflicting views, including that of experts. Data might be met with data; it might also be met with a professional critique. And of course it is possible that an individual expert will be able to

---

23 See Kahneman and Lovallo, supra note.
show, persuasively, that data does not establish what it claims. These sorts of disputes can be handled in the standard fashion. What we are suggesting here is that because individual experts are distinctly prone to error, it would be far better to begin the process with reliable evidence rather than particular recollection. If individual experts can show that the statistical data are wrong, the legal system will be better off for the demonstration. But in the long run, we predict, these demonstrations will be the exception rather than the rule.

B. Admissibility: Relevance and Hearsay

Statistical data are not typically used in negligence cases. In fact the legal system is uncomfortable with the use of such data. They might be excluded, as inadmissible, on several grounds.

It is possible to object that statistical data are not relevant, because they cannot determine the proper standard of care. In many states, there is a continuing dispute about the proper role of common practice in negligence cases. Where common practice is not determinative of the standard of care, the data cannot resolve the legal issue. Some courts have suggested that in such circumstances, the data should be inadmissible on grounds of irrelevance. But this is a mistake. Even when common practice is not determinative, it is pertinent, as reflected by the very fact that experts frequently are allowed to testify about what most doctors do. Statistical data should be found admissible for the same reason that expert testimony is admissible.

It might also be thought that statistical data should be seen as hearsay. Perhaps the witness seeking to introduce hearsay evidence is attempting to say what others have said, and perhaps these violate the prohibition on admitting materials of that sort. But this is a misreading of the hearsay rules. Under federal law, experts are allowed to rely on “statements contained in published treatises, periodicals, or pamphlets on a subject of history, medicine, or other science or art, established by the testimony or admission of the witness or by other testimony or by judicial notice.” Federal rules also allow experts to base opinions on facts and data of the kind reasonably used by experts in the field when they form opinions or inferences. Statistical data, if published, should be permitted under the first rule just quoted; if unpublished, they should be admissible to the extent that they are of the sort reasonably relied on by experts. The conclusion is that data should not be treated as inadmissible hearsay.

C. Textbooks?

Perhaps there are alternatives, other than reliance on statistical data, to the problem of idiosyncratic recollections of individual experts. One possible view is that the opinions of

26 See Schrag v. Chicago City, 265 Ill. 3d 338 *1914); People v., Anderson, 113 II 2d 1 (1986).
27 Fed. R. Evid. 803(18).
28 Fed. R. Evid. 703, 705. But see the puzzling ruling in Roach v. Springfield Clinic, 1992 Ill. LEXIS 204 (1992), refusing to allow references to treatises on the ground that their authors are not available for cross-examination. This seems to us an anachronistic and odd reading of the prohibition on hearsay.
experts, as codified in textbooks or journal publications, should define the standard of care. By this view, when actual practices differ from the theoretical or recommended standards, it is the practitioners who are at fault. They may not know the standard, may disagree with it, or have misinterpreted it, but the burden of proof falls on the practitioners.

We believe that this approach is oversimplified and misguided. Under the law, the ordinary practice is important for its own sake, whether or not it is decisive. Unless a textbook is actually based on ordinary practice, it will not report it reliably. If courts are going to consider the ordinary practice, they should have an accurate understanding of what it is.

Perhaps the textbook or journal can state a recommendation, one that is worth considering even if it does not capture ordinary practice. This is not implausible. But interpretation of the recommended standards (as expressed, for example, for AB-TIME in authoritative texts) is not always straightforward, and one can rarely find authoritative interpretations of authoritative texts. In the particular case of bacterial meningitis, what, in the minds of a lay jury, is a reasonable interpretation of textbook recommendations that antibiotics should be administered “promptly,” 29 “immediately,” 30 or “at the earliest possible time” 31 to children with suspected meningitis. In such a context is 30 minutes too long, - is one hour, six hours? The recommendation in one text that antibiotics be administered “within 30 minutes after the diagnosis of meningitis is established” 32 does little to resolve the question, as the issue promptly arises - when exactly is the diagnosis of meningitis established? If the question is what do doctors do, it would be far better to rely on actual data about physician performance.

**Conclusion**

We have suggested here that there is general reason to believe that expert judgments about the standard of medical care will be erroneous, and that the errors will run in a predictable direction. The usual reason is optimistic bias, as people tend to believe that things can be done more easily, more rapidly, and more successfully than the evidence suggests. To establish this claim, we have drawn on existing evidence, highly suggestive on this point, and more particular evidence, presented here, of mistaken reports with respect to standard of care.

Our principal innovation has been to suggest that in light of the evident mismatch between expert recollections and empirical reality, the legal system should rely, wherever it can, not on the former but on statistical evidence of the latter. The best reason for reliance on individual recollections has been an absence of statistical evidence; but this is a gap that is rapidly being filled, and that is likely, in the next generation, to be replaced with a great deal of reliable information. Our emphasis has been on the question of medical malpractice, but the general implication is far broader: In any case in which a disputed question calls for expert testimony about ordinary practice, it is hazardous to rely on what particular experts recall, and far more sensible to make the outcome turn on statistical evidence, at least if the goal is accuracy in adjudication.

29 See R.D. Feign et al., Textbook on Pediatric Infectious Disease (Saunders 1981).
Readers with comments should address them to:

Dr. William L. Meadow  
Department of Pediatrics  
University of Chicago Children’s Hospital  
C684, MC 6060  
5839 South Maryland Avenue  
Chicago, IL 60637  
WLM1@midway.uchicago.edu

Cass R. Sunstein  
Karl N. Llewellyn Distinguished Service Professor of Jurisprudence  
Law School and Department of Political Science  
University of Chicago  
1111 East 60th Street  
Chicago, IL 60637  
csunstein@midway.uchicago.edu
Chicago Working Papers in Law and Economics  
(Second Series)

13. J. Mark Ramseyer, Credibly Committing to Efficiency Wages: Cotton Spinning Cartels in Imperial Japan (March 1993).
34. J. Mark Ramseyer, Public Choice (November 1995).
47. John R. Lott, Jr. and Kermit Daniel, Term Limits and Electoral Competitiveness: Evidence from California’s State Legislative Races (May 1997).
60. John R. Lott, Jr., How Dramatically Did Women’s Suffrage Change the Size and Scope of Government? (September 1998)
64. John R. Lott, Jr., Public Schooling, Indoctrination, and Totalitarianism (December 1998)
67. Yannis Bakos, Erik Brynjolfsson, Douglas Lichtman, Shared Information Goods (February 1999)
68. Kenneth W. Dam, Intellectual Property and the Academic Enterprise (February 1999)
70. Cass R. Sunstein, Must Formalism Be Defended Empirically? (March 1999)
71. Jonathan M. Karpoff, John R. Lott, Jr., and Graeme Rankine, Environmental Violations, Legal Penalties, and Reputation Costs (March 1999)
75. Richard A. Epstein, Deconstructing Privacy: and Putting It Back Together Again (May 1999)
76. William M. Landes, Winning the Art Lottery: The Economic Returns to the Ganz Collection (May 1999)
77. Cass R. Sunstein, David Schkade, and Daniel Kahneman, Do People Want Optimal Deterrence? (June 1999)
78. Tomas J. Philipson and Richard A. Posner, The Long-Run Growth in Obesity as a Function of Technological Change (June 1999)
79. David A. Weisbach, Ironing Out the Flat Tax (August 1999)
81. David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages (September 1999)
82. Cass R. Sunstein, Nondelegation Canons (September 1999)
83. Richard A. Posner, The Theory and Practice of Citations Analysis, with Special Reference to Law and Economics (September 1999)
84. Randal C. Picker, Regulating Network Industries: A Look at Intel (October 1999)
90. David A. Weisbach, Should the Tax Law Require Current Accrual of Interest on Derivative Financial Instruments? (December 1999)
95. David Schkade, Cass R. Sunstein, Daniel Kahneman, Deliberating about Dollars: The Severity Shift (February 2000)
105. Jack Goldsmith and Alan Sykes, The Dormant Commerce Clause and the Internet (November 2000)