Empirical Evidence and Malpractice Litigation

Philip G. Peters Jr.
University of Missouri School of Law, petersp@missouri.edu

Follow this and additional works at: http://scholarship.law.missouri.edu/facpubs

Part of the Evidence Commons, Litigation Commons, and the Medical Jurisprudence Commons

Recommended Citation

This Article is brought to you for free and open access by University of Missouri School of Law Scholarship Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of University of Missouri School of Law Scholarship Repository.
EMPIRICAL EVIDENCE AND MALPRACTICE LITIGATION

Philip G. Peters, Jr.*

I. INTRODUCTION

Critics of medical malpractice litigation believe that expert testimony is often anecdotal and biased. To remedy this problem, several critics have recently suggested that courts “move toward greater reliance on actual data, and less reliance on the recollections of isolated experts.” Specifically, they suggest that social science methods be used to ascertain what physicians actually do in circumstances like those involved in the disputed case. With this evidence, proponents argue, juries would get a more accurate picture of clinical norms than they currently receive from the recollections and anecdotal experiences of individual expert witnesses.

Unquestionably, reliable empirical evidence of clinical practices has the potential to improve upon the adversarial descriptions of reality that juries now hear from hired experts. By improving upon the jury’s understanding of clinical reality, empirical research about clinical norms could lead to fairer verdicts.

However, the extent to which empirical evidence is likely to improve the system of justice should not be overstated. At least three factors could limit its impact. First, there are practical obstacles to the funding of this research. Second, some claimed negligence, such as the exertion of excessive pressure on the spinal column during surgery, will not be amenable to a quantitative study of customary practices under similar circumstances. Third, researchers will need to be especially careful to design unbiased

* Associate Dean and Ruth L. Hulston Professor of Law, University of Missouri-Columbia.


2. Meadow & Sunstein, supra note 1, at 641.
Nevertheless, these concerns should not obscure the concrete benefits that could be conferred upon malpractice litigation by more accurate evidence of clinical norms. Better empirical evidence will help courts reach more accurate verdicts. As a consequence, it will aid both deserving patients and wrongly-accused physicians. If our expectations are realistic and the design pitfalls are avoided, then the increased use of empirical research would constitute an important improvement in malpractice adjudication.

This Article is divided into four parts. Part II outlines the reasons why better malpractice evidence is needed. Part III describes some practical limits on this research that should temper our expectations. It also spotlights some design pitfalls that will have to be avoided. Part IV then explores how the data might help to decide actual cases. Finally, Part V explores in more detail the special problem posed when no medical consensus exists about the standard of care.

II. THE NEED FOR BETTER EVIDENCE

In most states, the legal standard of care for physicians is determined by the medical profession. In these states, proving the standard of care means proving what physicians customarily do under similar circumstances. Most of the remaining states use a reasonable physician standard that assigns the task of standard-setting to the jury rather than to the medical profession. In these states, evidence of customary practices, while no longer conclusive, is still admissible, highly relevant, and likely to be afforded great weight.

As a consequence, accurate evidence of clinical customs is crucial in malpractice litigation. Yet, critics believe that juries often receive unreliable evidence about medical customs from partisan experts hired by the parties. Experts have fallible memories and their experience is typically anecdotal and local, rather than systematic and comprehensive. David Eddy has described this problem as follows:

3. Philip G. Peters, Jr., The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium, 57 WASH. & LEE L. REV. 163, 163 (2001). This is just one respect in which physicians were the favorites of the courts during the early twentieth century. Id. at 192-201 (describing the laws and their gradual abandonment).

4. Sometimes, the standard is what physicians in similar localities would have done under the circumstances. BARRY R. FURROW ET AL., HEALTH LAW § 6-2 (2d ed. 2000).

5. Meadow & Sunstein, supra note 1, at 631.

In fact it is a major research task to figure out what practitioners in a community are doing. When an expert answers a question about a community standard it is extremely unlikely that he or she has any real data on actual practices. It is far more likely that what an expert believes is the practice in a community is what the expert believes should be the standard of care.\(^7\)

Note that Eddy is describing the difficulty of ascertaining local community norms, not national norms. Although the legal standard of care was once set by local custom, today nearly all states have moved to a standard of care based either on the customs in similar locations around the country or, in the case of specialists, to a national standard.\(^8\) Abandonment of the locality rule made sense as a matter of legal theory. However, the practical effect of the shift from local to national standards was to place formidable demands on expert witnesses. Physicians sparing a few hours of time from their practice to testify cannot realistically be expected to have accurate knowledge of the actual patterns of clinical practice around the country. As a result, many must generalize from their own (anecdotal) experience. Their testimony is, thus, much more likely to describe what they and their colleagues do than what most physicians actually do.\(^9\)

Like all humans, expert witnesses are also subject to unfounded optimism about how well they and their peers ordinarily perform.\(^10\) This may make them vulnerable to the tendency of making unfairly harsh judgments in hindsight.\(^11\) In addition, hired experts face the natural tendency to favor “their” side,\(^12\) especially if they depend upon expert witness fees for a significant part of their income.

The jury must then evaluate this flawed testimony. Because the jury has no independent basis for determining which of the two very different pictures of clinical reality painted by the opposing experts is correct, critics fear that the jury will make this choice based on the speaking ability of the experts or based on the jury's

---

8. See, e.g., William J. Curran et al., Health Care Law and Ethics 343-44 (5th ed. 1998); Furrow et al., supra note 4, at 271-73.
9. See Hartz et al., supra note 1, at 551 (reporting a strong correlation between physicians' self-reported conduct and criticism of the care provided).
10. Meadow et al., supra note 1, at 40 (demonstrating this bias in physician estimates of how much time expires before a patient is treated with antibiotics); see Meadow & Sunstein, supra note 1, at 639.
11. Meadow & Lantos, Expert Testimony, supra note 6, at 586.
sympathy for the plaintiff. Thus, critics feel that more systematic evidence of clinical practices is needed.

Although these concerns about jury manipulability are vastly overstated, the recommendation itself is a good one. On the one hand, a rich body of social science evidence strongly refutes the allegation of pro-plaintiff jury bias in medical malpractice cases. The available evidence is also reasonably comforting on the issue of jury capacity. On the other hand, better evidence would make the jury’s job much easier and, therefore, has the potential to improve the quality of its verdicts. Accordingly, we should look very closely at the proposal that courts and attorneys make greater use of empirical and social science research techniques to ascertain the applicable standard of care.

III. SOME LIMITS AND SOME PITFALLS

There are several reasons to temper the hope that better empirical evidence will transform malpractice litigation.

A. Who will fund the research?

1. Independent Foundations and Medical Organizations

In an ideal world, empirical research on clinical practices would be sponsored by “neutral” entities, such as non-profit foundations and medical organizations. The resulting data would then be available for use by either party in future malpractice litigation. This data would give physicians better guidance ex ante about the level of care required in states using a custom-based standard of care and it would improve the fairness of malpractice trials post hoc.

13. See id.; Meadow & Lantos, Expert Testimony, supra note 6, at 593.
16. Depending upon the jurisdiction, this would mean studying either (1) what physicians actually do in practice or (2) what physicians think a responsible doctor would do. See infra text accompanying notes 29-35.
In a world with scarce health care resources, however, these organizations may be reluctant to take significant funds away from research that has the potential to improve patient outcomes in order to fund research that studies current practice patterns. Proponents of this research will need to convince the funding organizations that this research has value outside of the courtroom, such as identifying the extent to which clinical practices have incorporated recent scientific insights or evaluated the success of recent educational efforts.

The incentive to fund these studies will also be reduced by the knowledge that they quickly will become dated. In addition, medical organizations may correctly perceive that their fingerprint on a study would reduce the weight assigned to it by juries. Thus, funding entities may limit their participation to: (1) clinical settings that frequently result in litigation; and (2) settings where the findings will have educational value independent of their use in litigation. In other cases, litigants probably will need to fund the studies themselves or hope that an existing data bank can be borrowed for this purpose, a subject explored in this Symposium by Professor Mark Hall and his colleagues.

2. Litigants

New empirical research is expensive. In addition, it takes some control of the evidence away from the trial attorneys. Under these circumstances, the suggestion that parties sponsor more empirical research is likely to be received coolly by attorneys for both parties. Most likely, this research will be reserved for the most serious cases and even then only undertaken when the party sponsoring the study is very confident that the findings will support her case. Furthermore, unfavorable findings need not (and will not) be revealed to the other party or shown to the jury unless the study was co-sponsored by the parties or ordered by the court.

The parties may also be discouraged by the likelihood that the jury will discount the evidence. Research undertaken at the behest of a party too often reflects that particular party's interests, as research sponsored by drug manufacturers has shown. As a result,

17. In addition, parties in disputed cases will claim that their case is materially different from the setting studied in a prior research effort and this charge will often be correct. Meadow and Lantos would put the burden of proof on the person who claims such an exception. Meadow & Lantos, Expert Testimony, supra note 6, at 589.
18. Id. at 592.
studies funded by a party may be viewed with skepticism by the court and the jury. 21

Moreover, the expense of social science evidence will exacerbate the high cost of litigating malpractice actions. Litigation costs in medical malpractice actions are already so expensive that only the most catastrophic medical injuries are worth litigating, no matter how compelling the evidence of carelessness. 22 With the added expense of social science research, even fewer worthy claimants would get their day in court. At the same time, added litigation costs will increase the financial pressure on blameless defendants to settle weak claims rather than litigate them.

To summarize, most of the empirical research on customs is likely to be performed by parties. Absent pressure from the trial court, a party is only likely to sponsor this research when the case has considerable settlement value and the party anticipates favorable findings. Given the impact of this research on litigation costs and its limited medical value, this relatively modest use of resources to fund litigation-sponsored customs research is probably appropriate.

3. Existing Databases

Others participating in this Symposium will explore the use of existing databases to ascertain clinical practice patterns. To the extent that this exploration bears fruit, proof of clinical norms could be cheaper and more credible than if it were funded by the parties for purposes of litigation. Thus, this is a very promising line of inquiry.

4. Courts

In the future, as courts learn that acquiring more accurate evidence of clinical norms is often feasible, they could insist that experts demonstrate a better evidentiary foundation for their testimony about customary practices than they have demanded in the past. They could, for example, use the spirit behind the Daubert line of cases to insist that experts in medical malpractice cases make some credible effort to sample industry practices before offering testimony about industry norms. 23 So far, the courts have not done


21. See Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1317 n.3 (9th Cir. 1995) (emphasizing the extra weight assigned to independent research).

22. I often hear about thresholds of $200,000-$300,000 in expected value.

so. However, it is not clear that defense attorneys have properly asked them to do so. My own sense is that defense lawyers have not settled on a strategy for capitalizing in malpractice cases on the broad judicial frustration with “junk science” and “hired gun” witnesses. One option would be to ask courts to raise the minimum level of preparation required before testifying about national clinical practices.

If the courts accept this invitation to make expert witness testimony more empirical and less anecdotal, they are likely to start by gradually increasing the homework that expert witnesses are required to do. The courts might, for example, insist on a better effort by expert witnesses to familiarize themselves with actual clinical norms, perhaps by undertaking at least a modest sampling of the relevant community. The jury would then determine the weight to assign to the testimony, depending on the quality of this expert’s inquiry. If the courts move in this direction, then attorneys hoping to maximize the credibility of their own experts would be encouraged to sponsor more reliable studies.

5. Dispute Managers, Including the Courts

Having an accurate picture of clinical norms would help mediators, arbitrators, and judges to foster settlements by allowing these dispute managers to give the parties an independent assessment of their case. An independent study of clinical practices commissioned by the dispute intermediary would act similarly to an appraisal in real estate disputes. However, this idea is too untested to assess its prospects.

B. Research Pitfalls and Limitations

1. Some malpractice settings will be more amenable to quantitative research than others

Some clinical norms will be easier to document than others. As Daniel Shuman has noted, this inquiry involves a decision of whether to take some weight of the evidence issues away from the jury. Daniel W. Shuman, Expertise in Law, Medicine, and Health Care, 26 J. HEALTH POL'Y & L. 267, 268, 272 (2001).

24. Shuman, supra note 23, at 280-81. But see Williams v. Hedican, 561 N.W.2d 817, 829-30 (Iowa 1997) (applying Daubert and permitting testimony about the need for treatment of pregnant women exposed to chicken pox to be offered by an expert whose theory, while not yet widely peer-reviewed, had recently been published in a medical journal).

25. Meadow & Lantos, Expert Testimony, supra note 6, at 593 (giving examples of practices that are amenable to study); see Hartz et al., supra note 1, at 554 (suggesting that surveys would work well for cases involving failure to consult other sources, refer to a specialist, or order diagnostic tests); Meadow & Lantos, Proactive, supra note 1, at 2-3 (suggesting that empirical evidence would be useful for the study of the number of premature babies not treated in
Retrospective chart reviews, for example, can measure only those factors methodically recorded in the chart. Also, some clinical dilemmas will occur too rarely to make the study of actual cases practical. In those circumstances, researchers will need to ask physicians what they think they would do in a certain situation, rather than observing what they actually do. Even if the possible biases associated with surveys can be avoided, some malpractice settings will be too complex or subtle for meaningful quantitative study. Furthermore, cases turning principally upon a factual dispute about the care actually provided or about its causative role will not be candidates for an empirical study.

2. Researchers will need to ask legally relevant questions

In their quest to improve malpractice adjudication, social science researchers will be entering an arcane world in which the legally important issues may not readily be apparent. The potential to take a wrong turn is illustrated by a study published in 1993. It was based on an actual case in which a baby had arrived at a hospital with an unknown malady. Later, a diagnosis of possible meningitis was made and antibiotics were initiated. Nevertheless, the child had a poor outcome. In the ensuing litigation, the child’s expert said that the standard of care was to start antibiotics within thirty minutes. The researchers had grave doubts about this testimony and set out to test it. They found that the median time for initiating antibiotics was 120 minutes from arrival, not thirty minutes.

It is easy to understand why the authors were troubled by testimony that they correctly sensed was erroneous. Nevertheless, their effort to discredit the expert witness resulted in the pursuit of a fact that was less important than others that they could have studied. The time from arrival to prescription should not have been a central issue in the case, as that time will understandably vary depending upon the patient’s symptoms at arrival. The truly relevant question is whether the child’s condition in the actual case dictated the diagnosis of meningitis at a time significantly earlier than the delivery room or sent home without apnea monitors).

26. See Hartz et al., supra note 1, at 547.
27. See id. at 553–54 (discussing hindsight bias and self-serving bias). For example, one of the shortcomings of surveys is that physicians may report what they think they should do, rather than what they actually do. See id. at 553 (noting that one study found that a survey using vignettes produced better indicators of actual behavior than chart review).
28. See id. at 547. Presumably, the feasibility will turn on whether the dispute can be distilled into two scenarios that could each be presented to respondents. In some of these cases, however, the standard will not be in dispute.
29. See Meadow et al., supra note 1, at 40; see also Meadow & Sunstein, supra note 1, at 637 n.33.
than the actual diagnosis. If not, then it really does not matter how long the child sat in the emergency room or whether that time exceeded the median. This is a point that the defendant's attorney and expert witnesses should have made clear to the court and jury.

3. **Surveys must be designed to minimize respondent bias and error**

Researchers must anticipate the risk that physicians responding to their surveys will not provide accurate answers or that they will not answer the question asked.\(^{30}\) Consider first surveys designed to find out what physicians actually do. Surveys, unlike chart reviews, introduce the risk that the respondents will provide answers that are self-serving or are otherwise incorrect. Physicians, for example, have been shown to be prone to an optimistic bias. The bias appears to affect their estimates about medical practices. In a study by Meadow and others, for example, physicians regularly overestimated how quickly antibiotics are usually prescribed for meningitis.\(^{31}\)

The same rose-colored glasses can also affect a physician's perception of her own conduct. For example, one study found that physicians significantly overestimate the amount of time they spend talking to their patients about healthy living and disease prevention.\(^{32}\) Thus, physician descriptions of their own conduct may be a less accurate picture of what they actually do and instead represent what they wish they did.

Surveys of physicians' practices can, therefore, incorporate their own systematic biases. Merely moving from the anecdotal recollections of a single expert witness (as is typical in today's litigation) to the collective recollections or opinions of a group of experts (as would occur using a survey) is no assurance that the information retrieved will be reliable. As a result, great care will have to be taken in the design and analysis of the first surveys in this field to determine the presence of bias and develop methods for avoiding it.

Similar confusion could confound surveys that seek to determine what responsible physicians *ought* to do (rather than what they actually do). Placing this issue to a vote, as surveys do, runs the risk of substituting majority practice for "good" practice or "best" practice. If, as some research suggests, respondents usually

---

30. The choice of whom to survey will also be very important. In one study, for example, the respondent's specialty significantly affected her answers. See Meadow et al., *supra* note 1, at 42 (finding that the estimates given by pediatric emergency experts were much more optimistic than those of specialists in pediatric infectious disease).
31. *See* Meadow & Sunstein,* supra* note 1, at 637-38.
believe that other physicians should do what the respondent currently does—even when the respondent’s practices fall outside of the range of acceptable practices—then the normative question will get a positive answer. In that event, evidence offered to the jury about what physicians feel is proper would simply be dressed up, uncritical evidence of what physicians think that they do. Although custom is a relevant barometer of what people in the field believe is proper, measuring custom is not the same thing as determining which customs are defensible. Whether surveys can help to educate the jury to distinguish good medicine from common medicine remains to be seen. One interesting strategy might be to ask physicians whether practices of practitioners adhering to a different school of thought also fall within the range of reasonable practice. Although a negative answer to this question ought not be definitive for obvious reasons, an affirmative answer would provide very powerful evidence that both schools of thought are defensible.

4. Obtaining sufficient detail

Researchers must also be careful to disclose the entire pattern of medical practice that they uncover and should not rely entirely on shorthand summaries, such as the use of a median or mean. In the study of meningitis medication times summarized above, the median time to medication was two hours. That information would have limited value if not accompanied (as it was in the study) by information about the distribution of actual times. The distributional data are needed because the median standing alone is consistent with a variety of practice patterns with very different legal implications. Tables I-IV, for example, all have the same median (120 minutes). Yet, they show very different clinical practice patterns.

33. See Hartz et al., supra note 1, at 551 (finding a strong correlation between self-reported practices and evaluations of the conduct of others). Most physicians assumed that other physicians would make the same decisions they did, even when their decisions were unusual. Id.

34. Although these Tables apply to a setting in which the choices are continuous, empirical evidence will also be useful when the choices are di- or trichotomous, as in choosing between two surgical techniques and a non-surgical intervention. See infra text accompanying note 40.
**Table I.**

<table>
<thead>
<tr>
<th>Minutes</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>

**Table II.**

<table>
<thead>
<tr>
<th>Minutes</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
</tbody>
</table>
### TABLE III.

<table>
<thead>
<tr>
<th># cases</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1 120</td>
</tr>
<tr>
<td></td>
<td>D2</td>
</tr>
<tr>
<td></td>
<td>D3</td>
</tr>
</tbody>
</table>

**median**

### TABLE IV.

<table>
<thead>
<tr>
<th># cases</th>
<th>Minutes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D1 120</td>
</tr>
<tr>
<td></td>
<td>D2</td>
</tr>
<tr>
<td></td>
<td>D3</td>
</tr>
</tbody>
</table>

**median**
As these Tables demonstrate, knowing that the median time to medication was two hours (120 minutes) would not tell us whether a genuine custom existed around that median (only Tables I and II show a coherent custom) and, if it did, how much variation around the median was routine (compare Tables I, II and IV). Indeed, a median of 120 minutes is equally consistent with both complete chaos (Table IV) and complete consensus (Table II).

IV. EVIDENCE OF A SUBSTANTIAL CONSENSUS

Notwithstanding the need for modest expectations about usage and for extreme care in study design, credible empirical evidence of clinical norms still has the potential to improve the fairness of malpractice litigation. Under either standard of care, reliable

---

35. Furthermore, use of a bare median to set the standard of care is also objectionable because it suggests that half of all the cases were treated negligently. Resisting that odd conclusion, courts have long held that the standard of care is set by the norms, not by the average. See, e.g., Nieves v. Hosp. Metropolitano, 998 F. Supp. 127, 137 (D.P.R. 1998); Morlino v. Med. Ctr. of Ocean County, 684 A.2d 944, 948 (N.J. Super. Ct. App. Div. 1996); Spray v. Bd. of Med. Exam’rs, 624 P.2d 125, 132 (Or. Ct. App. 1981). Once again, Tables III and IV demonstrate the point that it is really the range of reasonable variance around the mean that matters. As a result, many courts would rightly exclude evidence of the bare median or mean.

36. It could also produce some other, less central benefits. Here are five of them: (1) Maverick physicians would learn where they stand. Studies suggest
evidence that a substantial professional consensus exists will help
the jury to decide the case fairly and the parties to settle it
amicably. 37

Table I illustrates this point. The graph tells the jury that a
consensus does exist and where. By comparing the defendant’s
conduct to the graph, the jury can determine whether or not the
defendant’s conduct falls outside of the consensus. In this case, D1
clearly falls within the standard of care. Thus, the chart will help
D1 to defend against an expert who contends that a thirty-minute
delay is standard. 38 D3 has not complied with the standard of care

that outliers do not realize they are eccentric. Perhaps data showing their
eccentricity would prompt them to revisit their practices. (By contrast,
however, better evidence of practice patterns could magnify the pressure on
innovative physicians to stay in the pack.) (2) Reliable evidence of actual
practice patterns could help physicians or organizations that set very high
goals. The evidence of actual practices might help them to convince the jury that their
failure to meet their own goals was not a failure to provide competent care. (3) Reliable evidence that several schools of thought exist may help cost-conscious
physicians. For cost-conscious physicians, social science evidence showing that
many others share their approach could offer a non-inflamatory way to defend
their practices, i.e., without requiring a physician to openly ask the jury for the
latitude to be cost-conscious. I am personally troubled by this attempt to
“finesse” the jury, but I confess that the desire to do so is an understandable
response to recent (albeit early) evidence that juries punish businesses that
explicitly take costs into account. See Viscusi, Corporate Risk, supra note 15, at
557. (4) Illuminating jury norms. By tracking how juries have decided cases in
which they have heard detailed evidence about practice patterns, we may
acquire a useful glimpse into the minds of juries. We might learn, for example,
how they expect physicians (or any defendant) to behave under conditions of
uncertainty. We might see, for example, whether they are willing to forgive
variations based on uncertainty about the efficacy of two alternative
treatments, but not disagreements over the marginal value of very expensive
and only slightly effective treatments. (5) Changing judicial expectations about
minimally sufficient evidence. One of the lessons of this Symposium is that
courts ought to ask to see the evidence behind an expert’s opinion. Hopefully,
recognition of this fact will carry over into malpractice cases where no social
science studies are offered. In those cases, neither judges nor lawyers should
let an expert leave the witness chair without revealing the evidentiary basis for
her opinion.

37. This seems to be the use that Meadow and Lantos have in mind. See
Meadow & Lantos, Expert Testimony, supra note 6, at 592 (suggesting that the
evidence would only be dispositive in clear cases); Meadow & Lantos, Proactive,
supra note 1, at 2.

38. For this plaintiff to prevail, she will need to convince the jury that the
customary delay is inexcusable. In theory at least, she will only be permitted to
do this if she is in a state that has rejected the custom-based standard of care.
In reality, courts often let plaintiffs get to the jury when their experts have
offered an aspirational practice, rather than a customary practice, as the norm.
As Richard Lempert correctly points out, better empirical evidence of custom
may make it harder for the courts to permit such aspirational testimony or for
the court to give these cases to the jury. Richard Lempert, Following the Man
on the Clapham Omnibus: Social Science Evidence in Malpractice Litigation, 37
WAKE FOREST L. REV. 903 (2002). For those of us who think that customs
should not be conclusive, this would be an undesirable consequence of better
and should have a good explanation for her conduct, while D2's status is unclear.

One interesting consequence of displaying clinical norms in this graphical way is that it reveals the normative discretion that the jury possesses even when the jurisdiction uses a custom-based standard of care and reliable evidence of clinical customs has been presented. The jury must still decide how far down the sides of the bell curve physicians should be allowed to vary without incurring liability for any harm caused. (Thus, D2's liability is hard to predict.) This decision requires a normative, rather than an empirical, judgment and it is the traditional province of the jury to render these judgments. The courts have not discussed whether jurors should be given any guidelines for using this discretion, presumably because the parties have not yet offered evidence which poses the issue so patently. Factors that ought to be relevant to the jury's decision about the range of reasonable choice include the degree of scientific uncertainty facing the physician, the length of time that current best practices have been known, the limits on local resources, and the unique characteristics of the patient.

Now consider Table II. It shows the same median as Table I, but a much smaller range of divergence (i.e., a smaller standard deviation). The contrast between Tables I and II demonstrates how revealing empirical evidence can be. With a graph like this, the jury will have a much more accurate picture of clinical norms than it could hope to receive from individual experts. Here, the information provided by the graph about the uniformity of consensus (shown by the compactness of the curve) will help the jury to decide how wide a range of clinical responses might be reasonable. Note the implications for D2!

Recall the study of meningitis treatment. Assume that researchers had found a cluster of times around the two-hour mark, as shown in Table I. Evidence of this distribution might lead a jury to conclude that any time within that cluster (say between 100 to 140 minutes) was "customary" or "reasonable." By contrast, evidence that variation is minimal, as in Table II, might lead the jury to identify a narrower range of permitted practice (e.g., 110 to 130 minutes).

empirical evidence. In my view, the better solution to this problem is to change the standard of care, not to exclude empirical evidence. The jury should understand that it is being asked to buck existing clinical norms.

39. See Meadow & Lantos, Proactive, supra note 1, at 5. This discretion greatly complicates the task of experts who testify on the issue of causation, because they need to determine whether the defendant's unreasonable delay was harmful. To ascertain that, the experts need to know how long of a delay would be regarded as reasonable.

40. See id. at 2.
V. EVIDENCE OF MULTIPLE SCHOOLS OF THOUGHT

In the past few decades, medical researchers have learned that clinical practices vary dramatically and inexplicably. A number of studies, beginning with the classic work of John Wennberg, have demonstrated that physician practices vary widely, even within narrow geographic limits. In Vermont, for example, eight percent of the people in one community had their tonsils taken out while seventy percent of the residents of a different community had the surgery. In Iowa, the rate of prostate removal ranged from fifteen percent to sixty percent. Furthermore, a Medicare study found that procedure rates varied by more than three hundred percent for more than one-half of the procedures studied.

Many factors combine to produce this variation. The most important factor is uncertainty. This uncertainty arises out of the "bewildering variety of individual characteristics, histories, signs, symptoms, and behaviors" and the limited information about the efficacy and risks of possible treatments, both at the population level and at the level of these widely different patients. Patients vary in ways that resist standardization. This variation in patients is matched by a similar variety in possible therapeutic responses, each with its own mix of benefits, risks, and costs. Because of the resulting uncertainty, reasonable physicians may disagree about the best way to proceed.

41. Portions of this section are drawn from Peters, Jury, supra note 14. This variation may help to explain why defendants and plaintiffs are sometimes able to find experts to support weak cases. When clinical preferences vary widely, there is likely to be at least one physician who will (sincerely) say what the lawyer needs to hear.

42. See, e.g., John E. Wennberg, Dealing with Medical Malpractice Variations: A Proposal for Action, 3 HEALTH AFF. 6, 6-7 (1984).

43. Id. at 9.

44. See id.

45. CURRAN ET AL., supra note 8, at 36.

46. See Eddy, supra note 7, at 391-92; Meadow & Lantos, Expert Testimony, supra note 6, at 587.

47. Eddy, supra note 7, at 391.

48. See MARK A. HALL, MAKING MEDICAL SPENDING DECISIONS 84-88 (1997) (concluding that individual treatment decisions are complex and individualized); James A. Henderson, Jr. & John A. Siliciano, Universal Health Care and the Continued Reliance on Custom in Determining Medical Malpractice, 79 CORNELL L. REV. 1382, 1390 (1994) (concluding that the "highly differentiated nature of medical problems" is an obstacle to the formation of useful medical customs); Allan H. McCoid, The Care Required of Medical Practitioners, 12 VAND. L. REV. 549, 584 (1959) (stating that "there is no 'standard patient'").

49. Henderson & Siliciano, supra note 48, at 1390.

50. See, e.g., A. Russell Localio et al., Identifying Adverse Events Caused by Medical Care: Degree of Physician Agreement in a Retrospective Chart Review, 125 ANNALS INTERNAL MED. 457, 462 (1996) (finding only a slight correlation in the findings of independent reviewers).
Furthermore, physicians vary in their preferences and in their knowledge of the medical literature. Often the science itself is uncertain, produced by gaps in the data or competing paradigms. Finally, the movement of many employers away from fee-for-service health plans and towards managed care plans has produced significant differences among health plans in their resources and their cost-containment philosophies. Under these circumstances, there will rarely be a "custom" that provides a clear rule of decision.

However, the range of reasonable choices is not unbounded. Some variations will not be defensible, including deviations caused by ignorance, inflexibility, habit, or obsolete information. Practices driven by the desire to maximize income may also fall outside the range of reason. David Eddy makes this point forcefully:

Indeed, given the very high rates of inappropriate care that can prevail in communities, if we actually measured what practitioners were doing and used that to define the standard of care, we would run a high risk of installing an inappropriate practice as the standard of care. The well-documented overuses of hysterectomies, antibiotics, bypasses, and C-sections are examples.

A. Potential Benefits of Empirical Evidence

Empirical evidence that uncovers a pattern of variance can provide three important benefits for the courts and attorneys who must evaluate malpractice actions. First, accurate empirical evidence will aid the court in determining whether the defendant's conduct fell outside of all accepted schools of thought. In these circumstances, sound evidence of the range of accepted practices will help the plaintiff prove her case and help the jury decide it.

Second, reliable evidence of multiple standards will prevent the jury from accepting erroneous claims by the opposing experts that a single standard of care exists and that it favors them. Under current practice, both experts try to "cull a single standard of care from the cacophony of opinion." By refuting the incorrect contention that a single standard exists, the empirical evidence will

51. Henderson & Siliciano, supra note 48, at 1391.
52. Id. at 1399-1400; E. Haavi Morreim, Medicine Meets Resource Limits: Restructuring the Legal Standard of Care, 59 U. Pitt. L. Rev. 1, 8-20, 94-95 (1997); see also Page Keeton, Medical Negligence—The Standard of Care, 10 Tex. Tech L. Rev. 351, 365, 368 (1979) (characterizing customs as "fictitious").
53. See Hall, supra note 48, at 103 (noting that physicians, when placed under scrutiny, may attribute more uncertainty to their work than actually exists).
54. Eddy, supra note 7, at 396.
transform the jury's inquiry away from a search for "the real" standard of care into a reflection on how to decide cases in which no single standard of care exists. That shift is highly important and is discussed further below.

Third, in a setting where more than one school of thought is likely to be found, researchers can specifically ask respondents whether they consider more than one school of thought to be reasonable. An affirmative answer to that question would provide very helpful information for the jury in its task of determining whether to treat both schools of thought as defensible. At the same time, highlighting the presence of two schools of thought will sharpen the jury's inquiry into the adequacy of the disclosure that preceded the patient's consent to one of those permissible approaches.

Finally, by making the ubiquity of medical variation more visible to the courts, empirical evidence may prompt the courts to clarify the rules governing the resolution of cases when there is no single standard of care.

B. Legal Issues Raised by Variation

For the courts, the mixture of excusable and inexcusable variation presents a dilemma. To do justice to both parties in the face of this variation, courts will need a standard for sorting good medicine from bad. They must fashion the law of malpractice so that inexcusable variations lead to liability, but excusable variations do not.

General negligence law makes no special accommodation for the problems raised when uncertainty produces a variety of responses, many of which may be reasonable. However, malpractice law already has a rule designed specifically for this situation. The "respectable minority" or "two schools of thought" rule permits physicians to choose among respectable schools of medical thought without fear of liability. This rule arises out of judicial unwillingness to choose among conflicting schools of thought when physicians themselves cannot reach a consensus. As a result,

56. See supra notes 36-54 and accompanying text.
58. See, e.g., Downer v. Veilleux, 322 A.2d 82, 87 (Me. 1974) ("[A] physician does not incur liability merely by electing to pursue one of several recognized courses of treatment."); Haase v. Garfinkel, 418 S.W.2d 108, 114 (Mo. 1967) (stating there can be differences of opinion among competent physicians); FURROW ET AL., supra note 4, at 382-84 (describing the respectable minority defense).
59. RESTATEMENT (SECOND) OF TORTS § 299A cmt. f (1965) ("The law cannot undertake to decide technical questions of proper practice over which experts
physicians are theoretically insulated from liability both when they comply with an established custom and when respectable medical opinion is divided. In theory, the respectable minority rule eliminates the need to identify a single professional norm.

However, courts have not applied the respectable minority rule in this fashion. Instead, courts give these cases to the jury. As Professor Mark Hall has noted, "[t]his breakdown between theory and practice essentially allows the jury to impose, based on its own independent judgment, the governing standard of care—the very result malpractice law attempts to avoid." In practice, therefore, the custom-based standard of care routinely fails to provide an external benchmark to displace the jury's independent determination of reasonability. Thus, the jury uses its best judgment just as it would in states that have abandoned the custom-based standard of care.

What does all of this have to do with the use of empirical evidence to prove clinical practices? First, empirical evidence will let the jury know when no single "gold standard" exists in clinical practice. This information will refashion the jury's task in the case. Its job will no longer be to determine which expert is most accurately describing the standard of care. Instead, its project will be to determine whether the defendant's school of thought is "respectable," i.e., whether it has a justifiable explanation, such as clinical uncertainty, or whether the defendant occupies a pocket of bad medicine prompted by habit or self-interest.

Second, frequent reliance on social science evidence will educate the courts about the frequency with which variation in practice occurs. This variation calls into question an underlying premise of the custom-based standard of care. By calling attention to the reality, ubiquity, and inexplicability (even to medical experts) of clinical practice variation, social science evidence may prompt those courts which still use a custom-based standard of care to rethink reasonably disagree.").

60. Mark A. Hall, The Defensive Effect of Medical Practice Policies in Malpractice Litigation, 54 LAW & CONTEMP. PROBS. 119, 129 (1991); see Peters, Jury, supra note 14, at 947 (noting the failure of the respectable minority rule to keep these cases from the jury).

61. One potential solution would be to make the respectable minority rule more concrete by using a strictly quantitative definition of respectability. However, that would employ a criterion that is a poor proxy for reasonability.

62. However, reasonability states have more appropriate evidentiary rules. To inform the jury, the parties can introduce any relevant evidence bearing on the reasonability of the defendant's conduct, including evidence of clinical practices and evidence of the efficacy, risks, and costs of the options facing the defendant. The defendant can offer evidence about any scientific uncertainties that complicated her decision. The plaintiff can introduce evidence showing that a particular practice is no longer a good practice, including journal articles, recent texts, evidence-based clinical practice guidelines, and the testimony of respected experts.
their deference to custom, at least when no dominant custom exists. That step would make the law more coherent and juries could then focus on the task of identifying the range of reasonable clinical choice.

C. Empirical Evidence and the Evaluation of Multiple Schools of Thought

At the outset, reliable evidence of practice patterns will help the court to determine whether two distinct schools of thought actually exist. On the facts depicted in Tables I and II, for example, a defendant whose expert attempted to describe the defendant's four-hour delay as consistent with one accepted school of thought would be revealed as an outlier. If, on the other hand, the study revealed a pattern like that in Table III, then the defendant's contention would be confirmed. The jury could then proceed to determine whether this school of thought was "respectable" or "reasonable." Evidence that the profession was evenly divided, as shown in Table III, would make the plaintiff's case very difficult. The plaintiff would need to have convincing evidence that one approach was obsolete or self-interested.

In Table V, by contrast, the defendant's school of thought is much smaller and less aggressive than its alternative. Here, the jury would probably expect the defendant, rather than the plaintiff, to offer a convincing explanation for adopting a less aggressive approach, such as recent evidence that premature prescription causes needless side effects or is ineffective. Table VI illustrates how empirical studies could provide similar information when physicians face a choice among non-continuous options, such as choosing between surgery and bed-rest.

VI. CONCLUSION

Notwithstanding its limitations, good empirical evidence of actual medical practices has the potential to aid the jury considerably. At least in theory, it could be useful in both the "easy" cases (where it reveals that a consensus standard of care exists) and also some of the harder cases (where clinical practices vary). At the same time, the evidence will make the phenomenon of medical variation more visible to judges, highlighting the need to give juries better guidance on issues such as permissible variation around a single norm and the identification of unacceptable schools of thought. Ideally, the new evidence of inexplicable variation in

---

63. For an extended discussion of the choice between these two standards of care, see Peters, Jury, supra note 14.
64. Patients can identify contexts in which self-interested overtreatment or selection of treatment is especially likely by using databases such as Dartmouth's Atlas of Health Care Database. See Wennberg & Peters, supra note 57.
clinical practice will also encourage courts to abandon their conclusive deference to medical custom in favor of a standard of care that makes custom (when it exists) relevant, but not dispositive.