

# Solving the Medical Malpractice Crisis

## *Use a Clear and Convincing Evidence Standard*

Eli Engel, MD, PhD; Edward H. Livingston, MD

**T**he medical malpractice crisis has smoldered for many years with few new ideas regarding how to improve matters. Physicians promote limits on plaintiff noneconomic damages, but this has been ferociously resisted by the legal community. They argue that limiting remuneration to patients harmed by negligent practices is fundamentally wrong. We hypothesize that malpractice litigation is out of control because of an excessively lax evidence standard. Raising the evidence standard from the current “more likely than not” to “clear and convincing” would sharply reduce medical malpractice judgments against physicians. Clear and convincing is an evidence standard currently in use by courts for certain cases, and its adoption for malpractice litigation would not limit compensation for injuries resulting from negligent practices and should be well received by the legal community. *Arch Surg.* 2010;145(3):296-300

Physicians consider the current state of the medical malpractice problem a crisis.<sup>1</sup> In response, several tort reform initiatives have been proposed with varying degrees of success. The predominant approach and the most highly publicized efforts for medicolegal tort reform rely on controlling the physician’s economic consequences of an adverse legal decision after the decision has been made. The medical community has made only modest progress in gaining public support for this sort of tort reform. This failure suggests that we have selected the wrong battleground. The Health Care Quality Improvement Act of 1986 granted physicians unprecedented legal protections when reviewing substandard practices during peer review processes. Immunizing physicians from bearing responsibility for errors they committed is viewed as unacceptable by many in the legal community.<sup>2,3</sup> The public perceives physicians and their insurers as a wealthy class and has little sympathy for their economic plight. If a physician is found guilty

of malpractice, why should limitations on financial penalties be imposed?

### *See Invited Critique at end of article*

Current tort reform proposals rely on damage control after a malpractice verdict is rendered. Rather than wait for an adverse outcome, attention should be directed to the courtroom process before decisions are made. Medical malpractice is defined as care delivered by a physician that falls below the community standard. During a trial, a jury hears evidence regarding the standard of care and then applies certain rules for evaluating the evidence to reach a decision. Various types of legal proceedings rely on different evidentiary standards. The most rigid evidence standard is “beyond a reasonable doubt” and is applied to criminal cases. “Clear and convincing” is an intermediate standard and is applied to custody decisions and medical board actions. The most liberal evidentiary standard is the “preponderance of the evidence,” which translates to the simpler expression “more likely than not.” It is applied to tort cases such as medical malpractice.

**Author Affiliations:** Department of Biomathematics, University of California, Los Angeles (Dr Engel); Division of Gastrointestinal and Endocrine Surgery, The University of Texas Southwestern Medical Center, and Department of Surgery, Veterans Affairs Medical Center, Dallas, Texas (Dr Livingston); and Department of Bioengineering, The University of Texas, Arlington (Dr Livingston).

We propose that simple conversion of the tort evidence standard from more likely than not to clear and convincing will solve most of the medical malpractice problems for physicians in a way that is palatable to the public and to the legal community. We also demonstrate sound scientific principles on which this evidentiary standard is based.

#### HOW EVIDENCE UTILIZATION CHANGES THE PROFESSIONS OF LAW AND MEDICINE

In medicine and law, processes exist for quantifying the strength of evidence and making decisions based on these assessments. The history of how these concepts developed and the resultant systems for evaluating evidence is widely divergent between the two fields.

Scientific rigor applied to medical decision making is a relatively recent phenomenon. In the 19th century, medical education was informal and there were no established standards for accepting scientific evidence. Change was driven by the recognition through formal clinical study that therapies such as phlebotomy and blistering were of no value. In the same epoch, antiseptic surgery and vaccination were shown to be effective interventions through formal clinical trials. Demonstrating that the scientific method led to better therapies resulted in a scientific approach to medical education and practice with an emphasis on teaching scientific principles rather than having medical students simply memorize information, as was the emphasis in the pre-Flexner medical schools.<sup>4</sup> In the post-Flexner era, medical schools became major research institutions, leading the way to scientific inquiry and improvements in medical care driven by solid research programs.

Statistical tools used to assess scientific evidence were developed at about the same time medical schools were restructured. The theoretical basis for understanding small-sample statistics was laid down by Student and Ronald Aylmer Fisher in the second and third decades of the 20th century.<sup>5-7</sup> From these early works comes the familiar Student *t* test, the concept of null hypothesis testing with the establishment of a  $P < .05$  for accepting scientific conclusions. Statistics provide a theoretical framework in which to accept or reject scientific findings in an objective and consistent way.

Progress in medicine has been astonishing during the past 80 years. Rapid advances have occurred because medical science assumes that current clinical concepts are primitive and in need of improvement. Change comes about because new ideas are tested through scientific trials. New concepts that attain the threshold of being scientific truths by virtue of statistical analysis of the evidence become accepted and are applied to medical practice. Medical progress is an evolutionary process and, because substantive change in the treatment of human patients will result from scientific study, the standards for accepting scientific evidence are very rigid.

Changes in the legal profession are resisted, with the process and motivation for change differing substantially from those of medicine. Legal sophistication has existed for hundreds of years and evolves slowly. The law values consistency over outcomes because outcomes as-

essment of legal principles is not ingrained into the legal culture. Fear of instability resulted in adoption of governmental systems that are slow to make major changes. As long as our society is stable, we assume that the law works.

Although it does so slowly, the law changes to address societal inequalities. The evidentiary standard beyond a reasonable doubt was developed in recognition of the serious consequences to an individual if found guilty of a crime. Jurors passing judgment would have to live with a clear conscience after making a decision resulting in the punishment of the accused. Fearing consequences from a Christian God or a threat to their afterlife, jurors were instructed to evaluate evidence beyond a reasonable doubt. Rather than a sense that the evidence itself was compelling, the jurors were comfortable in their assessment of the case based on their own moral certainty that they had made the right decision. The jurors would have to live with the consequences they might suffer if their decision wrongly harmed another person.

Medicine and law rely on evidence evaluation for their respective systems to function. The law seeks stability, whereas medicine requires frequent and rapid change and, consequently, the 2 professions have developed different standards for evidence evaluation. Both systems rely on the probabilities that an event happens or has happened and use different probability thresholds for accepting something as "true."

#### PROBABILITY THRESHOLDS IN MEDICINE AND LAW: THEORETICAL BACKGROUND

Double-blind, randomized controlled trials provide the most compelling evidence that ultimately changes medical practice. A common use is comparing the efficacy of 2 drugs for treating a particular disease. The null hypothesis, that is, the one to be nullified by the statistical test, is that both drugs are equally effective. After the trial is completed, the results are tabulated and evaluated by an appropriate statistical test. A *P* value is assigned.  $P < .05$  means that the probability that the results come from simple chance variation alone is less than .05. The stringent evidentiary standard of having only a 5% chance of being wrong is used for all medical research. This same standard is applied irrespective of the consequences of the proposed change in medical treatments. For example, the  $P < .05$  standard is used for the evaluation of drugs to treat cancer as well as for relatively low-risk situations such as the evaluation of treatments of allergic rhinitis.

Probabilities are also used in legal decision making but in a different way than they are applied to medical research. For most cases, the assignment of probabilities is based on qualitative and not quantitative information. Although medicine applies the single 5% standard of evidence for all types of research, both high and low risk, the law uses various degrees of probability that are proportional to the consequences of a wrong decision that might occur in a legal proceeding.

The prosecutor in a criminal case or the plaintiff's attorney in a tort action has the responsibility of present-

**Table 1. Application of the Sensitivity, Specificity, and Predictive Value Table to Legal Proceedings**

Verdict	Actual Guilt		Predictive Values
	Yes	No	
Guilty	TP	FP	PPV = TP / (TP + FP)
Innocent	FN	TN	NPV = TN / (FN + TN)
	Sensitivity = TP / (TP + FN)	Specificity = TN / (FP + TN)	

Abbreviations: FN, false negative (a guilty person is found innocent at trial); FP, false positive (an innocent person is found guilty at trial); NPV, negative predictive value; PPV, positive predictive value; TN, true negative (an innocent person is found innocent at trial); TP, true positive (a guilty person is found guilty at trial).

ing information known as the “burden of proof,” which allows a jury to assess the evidence. This burden consists of 2 components. The first is the definition of the legal framework for the trial. The second is persuasion of the jury to agree with his or her arguments to a particular degree of probability, the level of which is included in the jury instructions. A lighter burden is required in cases considered to be of lesser importance. In a tort trial, the plaintiff’s attorney has to show only that the case is more likely than not to be true. This was originally termed *preponderance of the evidence*. In sports, this would be akin to a margin of 1 point in triple overtime.

When the ramifications of erroneously punishing a defendant are great, the evidence standard is very high. In a criminal jury trial, the standard is beyond a reasonable doubt. Criminal cases require the unanimous vote of a jury of 12 as opposed to tort trials, in which a vote of 9 of 12 is sufficient. This very high standard minimizes the risk of falsely imprisoning someone for a crime they did not commit. The trade-off, of course, is that some persons who are guilty will go free.

A third standard has arisen somewhat spontaneously in the United States during the past century. The standard of clear and convincing has emanated from common law traditions to deal with situations in which the existing 2 standards were deemed too lax or too strict. The courts define clear and convincing as intermediate between more likely than not and beyond a reasonable doubt.

The courts do not accept assigning *P* values to any of these criteria. However, it is conceptually useful to view the legal standards of evidence in the familiar terms of *P* values. For example, beyond a reasonable doubt would be equivalent to assigning a very small *P* value. A preponderance of evidence is defined as “50% plus a feather” or “a scintilla of difference.” This translates to a false-positive rate of less than 0.5, and this is satisfied by  $\kappa=0.49$ . Clear and convincing would be somewhere in between these extremes.

#### APPLICATION OF PROBABILITY THEORY TO THE LAW

When a person goes to court, he or she is either guilty or innocent of the accusation. The trial serves as a test

of those possibilities. Like a laboratory test, the courtroom process culminates in a verdict that may or may not be correct. Ideally, we would like to have a courtroom process and rules of evidence that, in general, would successfully differentiate between the guilty and the innocent. This is similar to a clinical test for a disease that has a 95% sensitivity and specificity. **Table 1** extrapolates the familiar sensitivity/specificity table used for clinical trials into a legal framework. The columns usually labeled as having or not having a disease are replaced with columns concerning actual guilt, and the rows associated with positive predictive values (PPVs) are replaced with rows concerning the verdict.

In clinical medicine, it is well known that a test with a very high degree of sensitivity (the ability to correctly establish a diagnosis in someone with a disease) and specificity (the ability to rule out a disease with a negative test result) may perform poorly if the disease is relatively rare. Diseases with low prevalence may have poor positive and negative predictive values. If one applies the test to an entire population, even if the test is very good at detecting the disease, the number of positive test results will be low compared with the total number of tests administered because the disease is relatively rare.

By extrapolating this example to legal proceedings, one can estimate the effect various evidentiary standards have on the incidence of false guilty verdicts against physicians and the effect that the prevalence of true malpractice has on these rates. We use this example to demonstrate that using the more stringent evidentiary standard of clear and convincing rather than the current preponderance of the evidence can reduce the number of false-positive courtroom judgments against physicians.

From this construct, the risks of falsely accusing someone in terms of the degree of evidentiary standards can be estimated. In the legal construct estimates of the PPV, that is, the probability that someone who is really guilty is given a guilty verdict, as a function of sensitivity, specificity and the prevalence of guilt in a population can be estimated. A sensitivity of 1.00 means that someone who is really guilty is always found guilty in court. When the sensitivity drops to 0.50, there is a 50/50 chance that someone who is truly guilty is found guilty in court.

A specificity of 0.99 means that someone who is innocent has a 99% chance of being found not guilty in court. Similarly, a specificity of 0.50 means that the person who is innocent has a 50/50 chance of being found not guilty in a trial. We have estimated that the clear and convincing evidence standard is equivalent to the 95% probability standard used in scientific inquiry. This standard is much more rigorous than the preponderance of evidence (50/50) but not as compelling as beyond a reasonable doubt (99%).

In **Table 2**, we calculated the PPVs for the prevalence of professional malpractice. We model malpractice prevalence ranging from 0.5% to 5%. Prevalence is defined as the proportion of adverse outcomes that is due to actual malpractice.

Table 2 emphasizes the high probability of finding someone guilty who is not really guilty if their offense is relatively rare given the current evidence standard of more likely than not. One can see that raising the specificity

**Table 2. Table of Positive Predictive Values for Malpractice Occurring at Differing Rates With Varying Sensitivity and Specificity<sup>a</sup>**

	Specificity				
	0.99	0.98	0.95	0.90	0.50
<b>Sensitivity = 1.00</b>					
Proportion of true malpractice					
0.005	0.330	0.200	0.091	0.048	0.010
0.01	0.500	0.340	0.170	0.092	0.020
0.02	0.670	0.510	0.290	0.170	0.039
0.05	0.840	0.720	0.510	0.340	0.095
<b>Sensitivity = 0.50</b>					
Proportion of true malpractice					
0.005	0.200	0.130	0.048	0.025	0.005
0.01	0.340	0.200	0.092	0.480	0.010
0.02	0.510	0.340	0.170	0.093	0.020
0.05	0.720	0.570	0.340	0.210	0.050

<sup>a</sup>Specificity columns correspond to the different evidence standards. A specificity of 0.99 corresponds to the “beyond a reasonable doubt” standard; a specificity of 0.95, to the false-positive rate of 0.05, which is considered acceptable in all scientific clinical trials; and a specificity of 0.50, to the “more likely than not” standard.

by raising the evidence standard substantially reduces the likelihood that someone will be falsely convicted. The probability that a bad event is actually due to true malpractice is low. Given that, the risk for false-positive verdicts is high.

The PPVs in Table 2 are very small numbers, especially in the more likely than not column. This indicates that, when the true negligence rate is low, guilty verdicts are associated with a large number of unjustified convictions because of the low specificity of this evidence standard. Raising the evidence standard to clear and convincing improves but does not eliminate this situation. Are the PPVs too low? Considering the rigors or professional training and certification, truly negligent acts occur infrequently. Undesirable outcomes and adverse events are relatively common, but negligence as their cause is relatively uncommon.

Several studies with remarkably similar results have established that there is an approximately 2.9% to 4.6% adverse event rate in hospitals. When these events were reviewed in the context of standards of care, approximately one-third were found to be due to negligent medical care. This translates to an overall incidence of 0.8% to 1.0% of true negligence in hospital admissions.<sup>8-10</sup> Overall, 2% of all negligent acts result in medical malpractice claims, but only 17% of all malpractice claims result from truly negligent activity.<sup>1</sup>

These statistics demonstrate that, although relatively few true cases of malpractice are litigated, most cases are without merit. Medical malpractice is a relatively rare event, occurring with an overall prevalence of about 1%.

Table 2 can be used to study the effect of differing evidentiary standards on medical malpractice litigation and trial outcomes. Assuming that there is perfect sensitivity for the trial, physicians who are guilty of medical malpractice go to trial and are never found innocent (sensitivity = 1.00). Because the incidence of malpractice is 1%, if there were 1000 trials, there would be 10 cases of true guilt, all of them found at trial. That leaves 990 cases in which no malpractice was committed. Using the preponderance of evidence standard, there is an approximately

50/50 chance that a jury will find a physician defendant guilty whether or not this is true. Thus, about 495 trials will find the innocent physician guilty and 495 will correctly find him or her innocent. This translates to a predictive value for a trial of 0.02 or only a 2% chance that the guilty verdict has identified a truly guilty physician. Increasing the evidence standard to clear and convincing will improve this rate to 17% because now only about 50 of 990 physicians who are not guilty will be wrongly found guilty at trial because of the evidence standard. Changing the evidence standard results in an 8- to 9-fold reduction in the number of trials wrongly finding innocent physicians guilty of medical malpractice.

In the real world of litigation, these numbers are somewhat distorted. Malpractice cases do not go to trial randomly. The high cost of legal process tends to discourage weaker cases. However, most malpractice suits, with their attendant legal defense costs,<sup>11</sup> are still filed against otherwise innocent physicians.

Numerous approaches to tort reform have been proposed.<sup>1</sup> That pursued most vigorously by physician groups has been the limitation of damage awards, a strategy resisted by the legal community. Our proposal to change the evidence standard from beyond a reasonable doubt to clear and convincing could substantially reduce the number of trials that wrongly find innocent physicians guilty without affecting a patient who has been truly harmed from appropriately recovering damages they deserve.

## CONCLUSIONS

The central and fatal error is to use a standard of proof that corresponds to a coin toss plus a scintilla. This is simply not valid under any circumstances. All other theoretical problems derive from this.

In tort actions, the State has no logical reason or compelling policy reason to favor the plaintiff or the defendant. The more likely than not standard yields the highest possible sensitivity and lowest possible specificity. This minimizes false-negative results but maximizes false-

positive results. This produces a major and completely unacceptable advantage of the plaintiff over the defendant.

Because most adverse results in medicine are not due to malpractice, a legal test must have high specificity to keep false-positive results at an acceptable level. The PPVs in Table 2 show this. The more likely than not standard fails woefully in this regard.

Our medical system is under extreme stress. Changing to a standard of clear and convincing would help to greatly reduce costs and redirect us toward better patient care and away from the practice of defensive medicine. This would be a simple action with zero cost. Although we have emphasized the potential impact of adopting a clear and convincing evidentiary standard for medical malpractice, this proposed change in the legal system would also benefit the outcome in tort cases relating to any type of professional malpractice, including all other health care professions, attorneys, accountants, architects, and engineers, to name a few.

**Accepted for Publication:** February 16, 2009.

**Correspondence:** Edward H. Livingston, MD, Division of Gastrointestinal and Endocrine Surgery, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Room E7-126, Dallas, TX 75390-9156 (edward.livingston@utsouthwestern.edu).

**Author Contributions:** *Study concept and design:* Engel and Livingston. *Acquisition of data:* Engel. *Analysis and*

*interpretation of data:* Engel. *Drafting of the manuscript:* Engel. *Critical revision of the manuscript for important intellectual content:* Engel and Livingston. *Statistical analysis:* Livingston.

**Financial Disclosure:** None reported.

## REFERENCES

1. Studdert DM, Mello MM, Brennan TA. Medical malpractice. *N Engl J Med*. 2004; 350(3):283-292.
2. Moore IN, Pichert JW, Hickson GB, Federspiel C, Blackford JU. Rethinking peer review: detecting and addressing medical malpractice claims risk. *Vanderbilt Law Rev*. 2006;59(4):1177-1206.
3. Scheutzow SO. State medical peer review: high cost but no benefit: is it time for a change? *Am J Law Med*. 1999;25(1):7-60.
4. Beck AH. The Flexner report and the standardization of American medical education. *JAMA*. 2004;291(17):2139-2140.
5. Livingston EH. Who was Student and why do we care so much about his t-test? *J Surg Res*. 2004;118(1):58-65.
6. Student. The probable error of a mean. *Biometrika*. 1908;6:1-25. doi:10.1093/biomet/6.1.1.
7. Kendall MG. Ronald Aylmer Fisher, 1890-1962. *Biometrika*. 1963;50(1-2):1-15.
8. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324(6):370-376.
9. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care*. 2000;38(3):261-271.
10. California Medical Association and California Hospital Association. *Report on the Medical Insurance Feasibility Study*. San Francisco, CA: Sutter Publications Inc; 2009.
11. Kakalik JS, Pace NM. *Costs and Compensation Paid in Tort Litigation: R-3391-ICJ*. Santa Monica, CA: RAND Institute for Civil Justice; 1986.

## INVITED CRITIQUE

# Solution to the Medical Malpractice Crisis: Don't Hold Your Breath

Engel and Livingston outline a rationale behind their novel proposal to change the current standard of evidence in a medical malpractice tort claim to a more stringent "clear and convincing" standard. Their reasoning and statistical data are clear and solid. The benefits of avoiding unnecessary malpractice litigation in time and cost and reducing the practice of defensive medicine are compelling.<sup>1</sup> Such articles describing problems with the American malpractice tort system are not new but stretch back to the mid-1800s.<sup>2,3</sup> Concerns regarding the potential legal (and medical) impact of quality improvement metrics, such as the Leapfrog standards, continue to the present.<sup>4</sup> Unfortunately, we may not see any real change soon.

Why doesn't the legal system change and adopt a new evidentiary standard that would benefit society as a whole? The answer lies in understanding the background of our legal system and how it retards rapid change, such as that proposed by Engel and Livingston.

There are 2 main versions of legal systems in the world: civil law (typically practiced in non-English-speaking countries) and common law (usually found in all English-

speaking countries). Civil law is based on laws that are codified by legislative bodies, and little latitude is given to judges to interpret those laws in individual cases. Civil law typically relies on career judges to render decisions and usually does not involve juries.

In contrast, common law is based on precedent, which is established by judges who interpret the facts of the case based on prior decisions. Common law typically relies on lengthy discussions of those interpretations, which then form the basis for new decisions in future cases. The finders of fact in common law systems are typically citizens who constitute jury panels and are instructed in their duties by the judge, who in turn relies on prior decisions of similar cases.<sup>5</sup>

Thus, change in the common law system comes primarily from slow incremental changes of society itself, through the collective judge and jury system rather than through legislative mandates in the civil law system. However, common law systems can indeed occasionally change in major ways by legislation (statutory law) or by elected officials (administrative law). Major change, for example, occurs by federal or state Supreme Court deci-