On Evidence, Medical and Legal

By Donald W. Miller, Jr., M.D.

Presented to the Seattle Surgical Society, May 22, 2006

Slide 1
Medicine, like law, is a pragmatic, probabilistic activity. Both disciplines require that *decisions* be made on the basis of available evidence within a limited time frame.

Slide 2: Decisions Made
Evidence is obtained in order to try and make the right decision. *Evidence* is information used to establish a fact or point in question.
In law judges and juries use evidence to determine responsibility and render justice. In Medicine, it is used to improve health and to make decisions on how best to prevent, diagnose, and treat disease. Evidence is marshaled to decide whether a research hypothesis is true or its null hypothesis is true—whether, for example, heparin causes thrombocytopenia or it does not do this.

Slide 3: Standards of Proof
Law has well-defined evidentiary standards. It sets different standards of proof according to the consequences of the decision, with life and liberty prized most highly. These standards range from the PP to the criminal standard. In most civil matters, evidence supporting or undermining a disputed proposition with a balance of probability greater than 51 %-- more likely than not--settles disputes and establishes liability. Evidence has to be "clear and convincing," however, to settle disputes involving child custody; involuntary commitment; withdrawal of life support; and to punish a person for a frame of mind driven by malice, oppression, or fraud. This standard is also used in administrative disciplinary proceedings against physicians and attorneys. The standard of proof used in criminal matters, "beyond a reasonable doubt," is sufficiently high that in 38 states it can result in the defendant being put to death. Science prizes objective certainty. It seeks evidence that is irrefutable and can withstand repeated tests of falsification. This standard of proof admits only a narrow range of evidence.

Slide 4: Evidence-Based Medicine
Evidence-based medicine aspires to a scientific standard of proof. It promotes the use of current best evidence in making decisions about the care of individual patients.
Its proponents festoon this definition with 3 adjectives, which are the "*conscientious*, *explicit* and *judicious*" use of current best evidence.

Slide 5: The EBM Evidence Pyramid
EBM proponents place well-designed randomized controlled trials and systematic reviews or meta-analyses of RTs at the top of the medical evidence pyramid, which they say meets the scientific standard of proof. Meta-analyses are better than a single randomized because they obtain greater statistical precision by combining multiple randomized trials that address similar questions.

Slide 6: RCTs on Transmyocardial laser revascularization (TMR)
Investigators have done seven randomized controlled trials on surgical transmyodardial laser revascularization, listed here.

Slide 7: TMR
In this procedure the surgeon burns 1-mm full-thickness holes through the left ventricle with a laser, 1 cm apart in a line from the base to the apex, and then in other lines 1 cm from each other, creating a total of 20 to 40 channels. These channels and the capillaries that grow out from them provide a way for oxygenated blood in the left ventricle to nourish the myocardium. The channels seal over on the epicardial side and mimic the sinusoids in a reptile's heart, which has no coronary arteries.

Slide 8-12: Results of the TMR trials
These randomized trials prove beyond a reasonable doubt, if not irrefutably, that TMR relieves angina better than medical treatment--and to a substantial degree, as this table shows--that it improves myocardial perfusion, whereas in the medically treated group perfusion decreases over a 12 month followup period--that patients undergoing TMR have a statistically significant better event-free survival--and an enhanced quality of life.

Slide 13: Biological Plausibility of TMR
Part of the factual matrix that places the benefits of TMR at the level of the scientific standard of proof is other supporting evidence that document its biological plausibility. Relief of angina does not simply result from denervation because the benefits of TMR extend out to 5 years, beyond the time when the nerves would have grown back. Its beneficial effect is a consequence of neoangiogenesis, resulting in improved myocardial blood flow. Accordingly, ACC/AHA guidelines now recommend TMR as a "Class IIA" therapy for intractable angina, which means the "weight of evidence is in favor of usefulness/efficacy" with a "Level of Evidence: A," meaning "data derived from multiple randomized clinical trials."

Slide 14: Patient Selection
The randomized trials on TMR are necessary to establish efficacy, but the guidelines derived from them do not provide information surgeons need to treat individual patients. Bypass grafts to 1 mm coronary arteries have a poor long-term patency rate. Should TMR be done also to prevent angina from occurring when these grafts occlude a year later? Some centers now do TMR routinely with reoperative bypass surgery. The mortality rate with TMR is high in pts. with poor LV function. Should the surgeon take on such cases using a balloon pump? To provide more specific information like this many more trials would have to be done in thousands of patients. Such is the case with other treatments. For example, Saver and Kalafut calculate that it would take 127 RTs in 63,000 pts. done over a 286-year period to determine the optimal combination of agents to treat Alzheimer's disease.

Slide 15: Epidemiological Evidence
Randomized trials provide epidemiologic evidence framed in terms of statistical significance. They address the incidence of disease and the effects of therapeutic interventions at the population level. They cannot detect rare events, and they cannot prove or disprove that *x* causes *y* in a specific individual. As the U.S. Federal Judicial Center's *Reference Manual on Scientific Evidence* states, "Epidemiology does not address the question of the cause of an

individual's disease. This question of specific causation is beyond the domain of the science of epidemiology." At best, it can establish a causal association, which is prone to two kinds of errors.

Slide 16: Bradford Hill Criteria

Epidemiologic evidence needs biological and other kinds of evidence to establish a causal association. The criteria listed here indicate what that evidence needs to be. Some U.S. courts, however, will admit epidemiology alone as evidence justifying an inference of causation in toxic tort litigation on a more likely than not basis when the relative risk is shown to exceed 2.0.

Slide 17: Verdict on the Null Hypothesis

Epidemiologic study commits a Type I error when the study incorrectly claims that the null hypothesis if false. An example of this type of error is a study published in the *Annals of Internal Medicine*, which concluded that Vitamin E supplements increase mortality, when, in fact, the null hypothesis, that Vitamin E supplements are safe, is true. A Type II error occurs when a study incorrectly claims that the null hypothesis is true. An example of this type of error is a study claiming that heparin does not cause thrombocytopenia, when the real truth is that in some patients heparin will destroy platelets and cause thrombotic complications, rendering the null hypothesis, that heparin doesn't do these things, false.

Slide 18: Cochrane Meta-Analysis on Albumin

Meta-analysis is the "gold standard" of evidence-based medicine. The Cochrane Group performed a M-A in 1998 of 30 randomized controlled trials on volume replacement in critically ill trauma pts. and found that the risk of death was 6 % higher in pts given albumin rather than crystalloid.

Slide 19: Cochrane Meta-Analysis on Albumin: The Fallout

When the study was published the London *Times* reported that it "suggests that up to 30,000 patients in Britain alone have died because they were treated with human albumin solution." Ian Chalmers, director of the Cochrane Centre in Oxford said that he would sue any doctor who gave him an infusion of albumin and that patients should seek redress in the courts for clinical negligence if the guidelines based on this analysis were transgressed.

Slide 20: Cochrane Meta-Analysis on Albumin: Its Flaws

This slide lists the flaws in this meta-analysis. Notably, none of the 7 authors of this study care for patients in the ICU; deaths < 24 hours after injury were excluded and >30 days included. And there was a major conflict of interest. Albumin is 30 times more expensive than crystalloid. The UK's National Health Service, which stocks albumin and crystalloid in its hospitals, funded the study, and it would stand to save a lot of money if it had to purchase only crystalloid and not albumin.

Slide 21: Cochrane Meta-Analysis on Albumin: Subsequent Developments

The Cochrane Injuries Group Albumin Reviewers updated their review in 2000 and did not alter their conclusion. Another systematic review on this subject was published a year later, which analyzed 55 randomized controlled trials, including ones that had a lower mortality with albumin that the Cochrane meta-analysis left out. This study concluded that albumin has no adverse effect on mortality. After this study was published the Cochrane folks then quietly removed their albumin review from its library of meta-analyses.

Slide 22: Scales for Assessing the Quality of RCTs
Analysts employ statistical techniques in their systematic reviews that include a numerical scale for weighting the quality of each trial. Juni and colleagues show how analysts can obtain diametrically opposing results depending on which of the more than 25 scales they use to distinguish between high- and low-quality randomized trials.

Slide 23: Does Low Molecular Weight Heparin Prevent DVT?
Quality scales divide trials into those deemed to be high quality or low quality. Analysts can pick the appropriate scale that provides the answer they want. This quality scale judges studies that show heparin has no effect of Deep Venous Thrombosis to be of low quality, whereas this scale does the opposite. Instead of meta-analysis being placed at the pinnacle of EBM, some observers contend that it is should be placed near the bottom of the pyramid as a form of "opinion-based medicine."

Slide 24: Evidentiary Flaws in Randomized Controlled Trials
Randomized trials and meta-analyses do not necessarily offer a scientific standard of proof because the evidence they supply is refutable. Biases in their methodology produce evidence that, in some cases, does not even meet the lowest legal standard of "more likely than not." These biases include faulty trial protocols, reporting outcomes in terms of relative risk without giving absolute risk of all-cause deaths, and justifying interventions on surrogate outcomes, like cholesterol level, when the more important primary outcome, freedom from myocardial infarction and survival, is not improved.

A study published in JAMA found that randomized trials funded by pharmaceutical companies are significantly more likely to recommend the experimental drug as the treatment of choice than studies funded by organizations that have no financial stake in the outcome
Chan and Altman reviewed 519 randomized controlled trials and found that incomplete reporting of outcomes (described in the methods section but not in the results section) was common. They conclude that the medical literature of randomized trials represents a selective and biased subset of study outcomes. As one observer puts it, "Epidemiological analysis is notoriously susceptible to misinterpretation, and even manipulation. Two sets of researchers can extract diametrically opposed results from the same data. The pharmaceutical and biotech industries now fund more than 60 percent of the randomized trials that medical journals publish, which raises the concern that supposedly objective science is being turned into a marketing tool. The editor of *Lancet*, Richard Horton, says "Journals have devolved into information laundering operations for the pharmaceutical industry." And a former editor of the *British Medical Journal* writes, "Medical journals are an extension of the marketing arm.

Slide 25: 2-Dimensional Guidelines
EMB issues guidelines for approaching clinical problems. These clinical practice directives are designed to treat single clinical problems. Derived from simplified clinical situations, they are not applicable to typical complex clinical situations.

Slide 26: Hypertension Guideline
The guideline for treating hypertension, shown here, is a standard 2-dimensional one. This one, released 3 years ago, has a new "prehypertension" level (120-39 systolic and 80-89 diastolic), which covers 22 % of American adults, 45 million people, who the guideline writers say should take the prescription drugs they recommend.

Slide 27: 3-Dimensional Guideline
What we, of course, see in practice are individual patients with multiple problems. Pts. with CAD often have other diseases like emphysema, hypertension, renal insufficiency, and sometime CLL, all of which have their own practice guidelines. Welsby describes the interaction of multiple diseases as one of 3-dimensional, Type II complexity. Skewed imposition of five interacting evidence based medicine guidelines applied to one patient makes it very difficult, if not impossible to decide, on the basis of guidelines alone, the best pathway to follow from symptoms and signs to therapy. William Osler had it right all along: "[Like law] Medicine is a science of uncertainty and an art of probability."

Slide 28: 4-Dimensional Guideline
Interaction between guidelines themselves results in an even greater 4-dimensional, Type III complexity. Indeed, the art of medicine lies within the matrix of interacting diseases and guidelines.

Slide 28: Cohort/Case Control Studies
From an Evidence-based medicine perspective, the least clinically relevant sources of information lie at the bottom of the EBM evidence pyramid and the most clinically relevant at the top. The four layers above case reports and case series represent actual clinical research, which include cohort and case control studies.

Slide 29: Observational Studies
Without randomization, statistical techniques are used to construct matched sets of Rx and control subjects, notably by constructing a propensity score and multivariate logistic regression modeling.

Slide 30: NEJM Aprotinin Study
This study, recently published in the NEJM, purports to show that aprotinin, used to reduce bleeding in heart surgery, blunt the systemic inflammatory response to surgery and cardiopulmonary bypass, and reduce the incidence of stroke, causes serious end organ damage. The day this article was published, the *NewYork Times* reported, "Compared with the other patients, those given aprotinin had twice the rate of kidney failure [5 percent]. They also had increases in other serious problems, including heart attacks, heart failure, strokes and a diffuse type of brain damage called encephalopathy." The report went on to say, reminiscent of the media response to the Cochrane albumin study, "Halting aprotinin use globally would prevent 10,000 to 11,000 cases of kidney failure a year and save more than $1 billion a year in dialysis costs, as well as nearly $250 million spent on the drug itself."

Slide 31: NEJM Aprotinin Study: Flaws
Aprotinin is a 100 times more expensive that aminocaproic acid, which is also used to reduce bleeding in heart surgery. The study was funded by a consortium of 160 medical centers that would stand to profit if they did not have to purchase aprotinin. The propensity score used to adjust for differences between Rx groups is prone to even more bias than quality scores used in meta-analyses. The study misapplied earlier work to their trial result, focusing only on result of aprotinin used with deep hypothermia and circulatory arrest
The authors

cited only 3 of 19 key references that support use of aprotinin for cardiac surgery, ignoring randomized trials that show aprotinin reduces stroke risk. And would the journal would have published this nonrandomized study if it had showed that aprotinin is better?

Slide 32: Indications for Aprotinin
Our current Indications for Aprotinin in Cardiac Surgery at the Seattle VA Medical Center, with input from cardiac anesthesia, are listed here.

Slide 33: Case Reports
Should case reports be this far down the evidence pyramid?

Slide 34: Value of Case Reports
The most essential evidence in medicine is the patient's story. In law, eyewitness testimony (i.e., a case report) can meet the highest legal standard of proof, of beyond a reasonable doubt. Medical evidence does not often meet the scientific standard of proof; and, as in law, it should be judged by standard of proof appropriate to the fact or point in question. An *anecdotal* case report can provide evidence of probative value, just like eyewitness testimony in a murder trial. And it can be similarly tested, by second opinions, re-examination, laboratory tests, and follow-up.

With *adverse drug reactions* a case report can surpass the highest legal standard of beyond a reasonable doubt and meet the scientific standard of irrefutability.

Slide 35: Double Hit Challenge-Dechallenge-Rechallenge (CDR) Evidence
Three events related to administration of a drug prove specific causation: 1) *challenge*—the adverse reaction occurs after the drug is given; 2) *dechallenge*—it resolves when the drug is discontinued; and 3) *rechallenge*—the adverse event recurs when the drug is given a second time.[43] Causation is judged to be certain owing to this "double hit," of challenge and rechallenge. Heparin causes thrombocytopenia in 2-3 percent of patients. In this patient the platelet count dropped from 200,000/mm3 to 60,000 after a 10-day course of heparin. Over the next 20 days, off heparin, it returned to normal (179,000). A second bolus of heparin was then given, which promptly dropped the platelet count to 49,000. No other causes for thrombocytopenia were evident, and the presence of heparin/platelet factor 4 antibodies provides biological plausibility for this reaction.

Slide 36: Causal Significance of CDR Evidence
The Institute of Medicine acknowledges the causal significance of CDR evidence in case reports, noting that the recurrence or non-recurrence of the adverse event—rechallenge—will have a major impact on causality assessment. Likewise, the FDA agrees and adds, "Assessment of temporal relationships and dechallenge/rechallenge information is usually considered your strongest evidence of a causal association. And the Stephens textbook states that a positive rechallenge is "probably the strongest proof of a causal relationship." If giving the drug a second time is not done owing to ethical considerations, then three cases of challenge-dechallenge only can prove causality, on at least a probable, more likely than not basis, if not beyond a reasonable doubt.

Slide 37: The Brides in Bath Case
The judiciary follows a set of rules on admissibility of evidence. With a few exceptions, hearsay evidence is not admissible, nor is opinions, except for expert opinion with regard to

technical and scientific matters. "Similar fact evidence" is admissible if it is relevant and probative, like in the *Brides in the Bath* case, where the defendant, George Smith, was accused of drowning his bride in the bathtub. No physical evidence implicated him in her death, but she had signed over her estate to him on their betrothal. Evidence was produced at trial that this person, using different names, had married two other women who also drowned in their bathtubs. They too had made financial arrangements from which he would benefit.
At the trial, the prosecutor had a nurse in a get in bath tub filled with water. He pushed and held her head under the water, and showed that she was unable struggle free and get out. In fact, he held her down too long and she almost drowned. This evidence was sufficient to find Smith guilty as charged, and he was executed (in 1915 in the UK).
Similar fact evidence like this in a legal setting is equivalent to CDR evidence in medicine. Both demonstrate causality at the high criminal standard of proof. In the brides in the bath case, their deaths precluded a dechallenge and a rechallenge, but such evidence nevertheless is essentially the same as three challenge-dechallenge cases in proving causation.

Slide 38: Autism
An epidemic of autism afflicts American children today. 50 years ago fewer than 1 in 10,000 children had this devastating malady, but today, with the prevalence now 1 in 166, one in every 68 American families has an autistic child. This graft shows the number of students in U.S. schools age 6-21 with autism. In 1991, there were 5,400. Eleven years later this number had increased more than 20-fold, to 118,000 children with autism in 2002.

Slide 39: MMR Vaccine as a Cause of Autism
This case, and others like him, provides strong evidence that the MMR vaccine, the live-virus measles, mumps, and rubella vaccine, causes autism. This child, the grandson of a pediatrician in Maine, had normal development up until 15 months of age when he had his MMR vaccination. He then regressed and became autistic. With behavioral therapy and biomedical treatments he experienced steady improvement until age 4 when he was given a MMR booster shot. He then regressed into a state of severe autism. The Autism Research Center here at the UW, analyzing home videos, has documented that some children with autism do develop normally for twelve to 24 months then regress and become austistic. Public health officials and their respective medical establishments in the United States and United Kingdom will not accept this kind of evidence with regard to vaccines, stating: "The weight of currently available scientific evidence does not support the hypothesis that vaccines cause autism." For them, only epidemiologic evidence is sufficiently "scientific."

Slide 40: Precautionary Principle Standard of Proof
The lowest standard of proof, which requires little or no evidence, is the PP.

Slide 41: Precautionary Principle
The most recent Wingspread Declaration states, "When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not established scientifically." Based on the maxim "better safe than sorry," this new standard of proof increasingly governs state regulatory policy and international environmental law. Following this principle, governments can implement policies and regulations based on what they think *might* cause harm, even if there is no evidence that a hazard exists. Regulators employ this Principle to reduce supposedly harmful CO2 emissions, ban DDT, and bar planting of genetically engineered crops.

Slide 42: Calamities Resulting from the Precautionary Principle
The benefits achieved from having banned DDT are disputed; and, not having access to this pesticide, 50 million people have died from DDT-preventable malaria. DDT had virtually eliminated malaria worldwide by 1970 before it was demonized and banished on questionable evidence. Today 400 million people in various parts of the world contract malaria and 1 to 2 million people die from this disease each year, most of them children in Africa. Likewise, regulators and environmental activists only consider the unproved risks high-yield genetically modified crops pose to the environment and people's health and do not address their potential benefits in combating malnutrition and starvation in Third World countries.
Had the PP been in place at the beginning of the last century, all the major advances in science and medicine over the last 100 years, like quantum physics and open heart surgery, would not have occurred. Implementing regulations based on the PP must be questioned.

Slide 43: The Problems of Evidence in Evidence-Based Medicine
Evidence-based medicine puts major constraints on the care of individual patients. The authors of this study put it this way, "Important single studies, particularly if not done as randomized trials, may be omitted from the authorized collection of "best available evidence." For example, when insulin first achieved a rapid reduction in diabetic acidosis and when penicillin first eradicated bacterial endocarditis, the results in both instances came from observational rather than RCT research, and each set of results was reported in a single study. Despite the extraordinary efficacy of both treatments and their dramatic impact in clinical practice, neither study, if newly reported today, would be included in the Cochrane collection of authoritative evidence. The threat of official [government], corporate, or private abuse will always remain whenever any collection of information has been prominently heralded as the 'best available evidence.' A new form of dogmatic authoritarianism may then be revived in modern medicine, but the pronouncements will come from Cochranian Oxford rather than Galenic Rome."

Slide 44: Conclusion
An attorney, Clifford Miller (no relation) in the UK, collaborated with me on this study. We conclude that medicine needs to develop a better understanding of the nature of evidence and of evidentiary proof by emulating law's approach to evidence. Law in turn needs a better understanding of the shortcomings of medicine's current approach to evidence.