CHEST
Point/Counterpoint Editorials
CHEST | Volume 133 | Number 5 | MAY 2008

Point: Evidence-Based Medicine Has a Sound Scientific Base

The scientific basis of evidence-based medicine (EBM) rests on three principles. First, systematic summaries of the highest quality available evidence should inform clinical decisions. Second, wise use of the literature requires a sophisticated hierarchy of evidence. Finally, evidence alone is never sufficient to make clinical decisions; rather, it requires trading off benefits and risks, inconvenience, and costs, and in doing so considering patients' values and preferences.

The Need for Evidence

Prior to the promulgation of EBM and the systematic reviews that lie at the heart of EBM, expert evidence reviews and opinions disseminated in narrative textbooks and review articles were often idiosyncratic and arbitrary. The resulting recommendations were inconsistent, often lagged behind the evidence, and were sometimes contrary to the evidence.

Consider, for example, the use of thrombolytic treatment of acute myocardial infarction (Fig 1, *top*).[1] By 1980, 23 randomized control trials (RCTs) including 5,767 patients had examined thrombolysis; and, had a metaanalysis been performed, the cumulative results would have demonstrated the effectiveness of thrombolysis ($p < 0.01$). Nevertheless, $> 40{,}000$ additional patients were enrolled in subsequent trials, of whom half did not receive life-prolonging thrombolytic therapy. During this period, expert authors of review articles published contradictory recommendations and consensus lagged a decade behind the evidence. The consequences included inefficient use of research resources, trial patients unnecessarily denied therapy, and a much larger number of patients presenting in routine clinical settings who did not receive thrombolysis. This example represents an error of omission; in other instances, including prophylactic lidocaine and calcium antagonists in myocardial infarction, positive expert recommendations contradicted negative results from RCTs (Fig 1, *center* and *bottom*).[1]

A Hierarchy of Evidence

Using the best evidence to make informed decisions requires rules for identifying that best evidence. Some have equated EBM with a naïve enthusiasm for RCTs, and a dismissal of other research designs. The approach of EBM is in fact far more sophisticated. As presented in the writings of the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) working group,[2] EBM defines quality of evidence as our confidence in the magnitude of effect estimates for a patient-important outcome. High-quality evidence provides robust results; low quality leaves uncertainty, and the likelihood that best estimates will change with newer, higher quality evidence.

Some forms of evidence suffer from a greater risk of bias than others; EBM suggests several hierarchies of evidence, specific for each type of clinical question. For example, high-quality studies of diagnostic tests require observational studies including an independent comparison of test results with a rigorous criterion standard. For questions related to prevention and treatment, the most rigorous research methodologies are N of 1 RCTs,[3] followed by multipatient RCTs, observational studies examining patient-important outcomes, physiologic studies, and unsystematic clinical observations.

Support for this hierarchy comes from empirical evidence that observational studies typically overestimate, but occasionally underestimate, the magnitude of treatment effects.[4] Table 1 presents examples of RCT evidence refuting results of observational or physiologic studies.

EBM recognizes that RCTs may sometimes provide only low-quality evidence. GRADE identifies five categories of limitations that may downgrade quality of evidence from RCTs.[2] First, methodologic limitations (including poorly concealed group allocation, lack of patient, clinician, or outcome assessor blinding, large loss to follow-up, or stopping early for efficacy) may bias study results. Second, small sample size with consequent wide confidence intervals may produce untrustworthy results. Third, RCTs may provide indirect evidence if the participants, interventions, comparators, or outcomes differ from those under consideration. For example, many trials measure

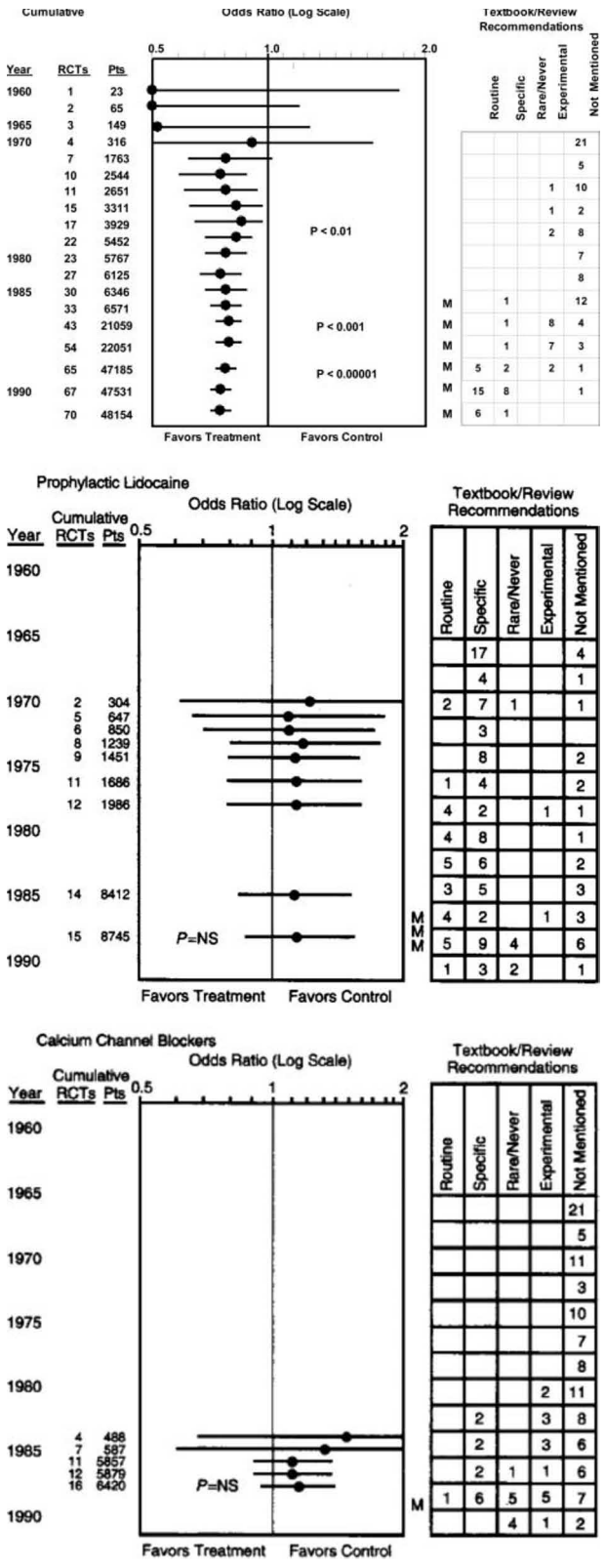

FIGURE 1. Cumulative metaanalyses and textbook recommendations. *Top*: Thrombolysis in patients with acute myocardial infarction. *Center*: Lidocaine in patients with acute myocardial infarction. *Bottom*: Calcium antagonists in patients with acute myocardial infarction. From Antman et al[1] (permission granted). Pts = patients; M = metaanalysis published; NS = not significant.

the impact of an intervention on surrogate outcomes, such as BP or FEV_1, but patients and clinicians are far more interested in outcomes such as mortality and quality of life. Fourth, inconsistent results leave us less certain. Fifth, selective publication of positive findings may bias results of systematic evidence summaries.

Furthermore, EBM recognizes that observational studies may provide high-quality evidence if treatment effects are sufficiently large, consistent and rapid. For example, no RCTs have demonstrated the impact of hip replacement in osteoarthritis on pain and function, but observational studies with huge, rapid effects leave us no doubt about the benefit of hip replacement.

Clinicians and patients must use the best evidence available (often observational or physiologic studies, or RCTs flawed by the five limitations listed above) to guide their decisions. EBM promotes understanding of evidence limitations, and the appropriate recognition of consequent uncertainty, in considering patient management decisions.

Evidence Is Not Enough

Critics of EBM contend that EBM fails to take into account the unique features of individual patients, and is applicable to population-level decision making but not to individual patient care.[20] This characterization could not be further from the truth: patients' values and preferences, and indeed, individual patients' decisions are central to the practice of EBM.[21]

EBM scholars have long recognized that patients' and clinicians' values may differ systematically. Consider, for example, treatment of patients in atrial fibrillation with anticoagulation to prevent strokes. Treatment with warfarin reduces the risk of stroke in these patients but increases the risk of serious GI bleeds. Traditionally, clinicians might have considered the best available evidence, and decided to administer an anticoagulant if they believed the benefits outweighed the risks, or elected to withhold treatment if they believed the risks were too great. Implicitly, this approach relies on clinicians' values and preferences.

Devereaux and colleagues[22] asked patients and physicians how many additional serious GI bleeds they would be willing to accept to prevent eight strokes—four minor and four major—in 100 patients. The results demonstrate that patients are far more stroke averse than clinicians, and that there is huge diversity in values and preferences among both patients and physicians (Fig 2).

The practitioner of EBM must consider the extent to which patients wish to be involved in decision making. For those who wish active involvement, they must communicate the expected effects of interventions in terms that patients can

Table 1—*Examples of Situations in Which Evidence From RCTs Refuted Evidence From Observational or Physiologic Studies

Question	Evidence From Non-RCT	Evidence From RCT
Does acetylcysteine prevent doxorubicin-induced acute myocardial morphologic damage?	An experiment in mice suggested that pretreatment with acetylcysteine 1 h before doxorubicin significantly decreased lethality, long-term mortality, loss in total body weight and heart weight.[6]	Acute doxorubicin-induced damage was similar in patients pretreated with acetylcysteine and placebo.[7]
Does naloxone improve neurologic outcomes in patients with spinal cord injury?	Naloxone significantly improved hypotension and neurologic recovery in cats subjected to cervical spinal trauma.[8]	No difference in neurologic outcomes, mortality, or morbidity in patients treated with naloxone or placebo.[9]
What impact do the antiarrhythmic drugs encainide and flecainide have on mortality from ventricular arrhythmias in patients after myocardial infarction?	A before/after study of patients with symptomatic, recurrent, previously drug-refractory ventricular tachycardia found encainide completely eliminated recurrence of ventricular tachycardia in 54% of patients for 6 mo of therapy. "Encainide is a safe, well-tolerated antiarrhythmic agent."[11]	Patients treated with encainide or flecainide had more than twice the risk (risk ratio, 2.64; 95% confidence interval, 1.60–4.36) of cardiac deaths and cardiac arrests than patients receiving placebo.[10]
In patients with acute lung injury or ARDS, what is the impact of inhaled nitric oxide on mortality?	Inhalation of nitric oxide in consecutive patients reduced the mean pulmonary artery pressure ($p = 0.008$), decreased intrapulmonary shunting ($p = 0.028$), and increased the ratio of the Pao_2 to fraction of inspired oxygen ($p = 0.008$).[12]	Inhaled nitric oxide did not increase the number of days patients were alive and off assisted breathing, nor alter mortality.[13]
In patients with severe emphysema, what is the impact of lung volume reduction surgery on mortality?	In consecutive patients, FEV_1 was significantly increased up to 36 mo after surgery ($p \leq 0.008$), the 6-min walk distance increased, and dyspnea improved after surgery.[14]	The 30-day mortality rate after surgery was 16% (95% confidence interval, 8.2 to 26.7%), as compared with a rate of 0% among 70 medically treated patients ($p < 0.001$).[15]
In patients in need of a pacemaker to correct symptomatic bradycardia, what impact do physiologic (AAI) and ventricular (VVI) pacing have on risks of cardiovascular morbidity and death?	In a cohort study, patients treated with VVI pacing had a higher incidence of congestive heart failure (37% vs 15%; risk ratio, 2.5; $p < 0.005$) and mortality (23% vs 8%; risk ratio, 2.9; $p < 0.05$) than patients treated with AAI.[16]	Type of pacemaker had no effect on the mortality or incidence of congestive heart failure. There were significantly more perioperative complications with AAI pacing than with VVI pacing (9.0% vs 3.8%, respectively; $p < 0.001$).[17]
In critically ill patients, what is the impact of treatment with growth hormone on mortality?	81% of patients who had failed standard ventilator weaning protocols and who were subsequently treated with human growth hormone were eventually weaned from mechanical ventilation with overall survival of 76% (significantly greater than the predicted survival rate; $p < 0.05$).[18]	Two RCTs demonstrated a higher mortality rate in patients receiving growth hormone relative to placebo (risk ratio, 1.9; 95% confidence interval, 1.3–2.9; and risk ratio, 2.4; 95% confidence interval, 1.6–3.5; $p < 0.001$ for both).[19]

*Adapted from Lacchetti et al[5] (permission granted). AAI = atrial single-chamber pacing; VVI = ventricular single-chamber pacing.

understand. Recently, investigators have developed "decision aids"—tools that present evidence in ways that patients understand—to help clinicians with the enormous challenge of effective communication of risks and benefits.[23] For patients reluctant to be involved, clinicians must explore their values and preferences and ensure that decisions are consistent with patients' worldview. Continuing technologic and methodologic advancement in decision aids and other tools to assist with knowledge translation should assist clinicians with the challenge of implementing EBM in practice.

Conclusion

Given the soundness of the three principles on which EBM rests, it is not surprising that guidelines based on these principles can improve outcomes. Consider for example, interventions targeted at decreasing catheter-related bloodstream infections in the ICU. High-quality evidence supports the effectiveness of hand washing, full-barrier precautions, skin cleansing with chlorhexidine, avoiding the femoral site if possible, and removing unnecessary catheters.[24] The introduction of an intervention to improve these practices in 108 ICUs resulted in a large decrease (up to 66%) in the rates of catheter-related bloodstream infection that was maintained throughout the 18-month study period.[25] Other studies have confirmed the positive impact that evidence-based guidelines may have on patient-important outcomes.

In summary, EBM represents a highly individualized, patient-centered approach to clinical decision making. The scope of EBM is evolving; with this evolution arise new tools to assist clinicians in the

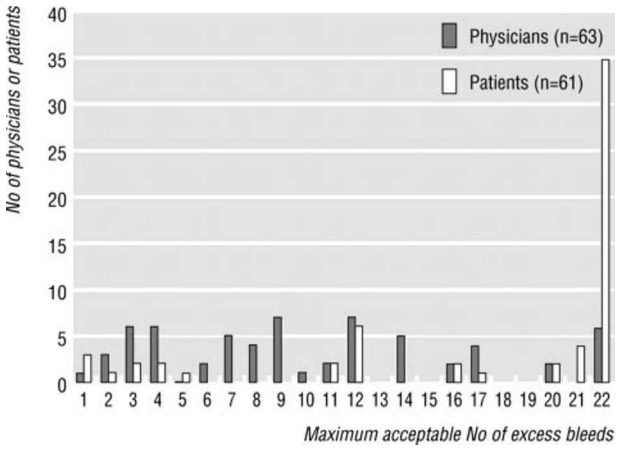

FIGURE 2. Varying thresholds of major GI bleeding in 100 patients found acceptable by patients and physicians when strokes are reduced by 8 in 100 patients. From Devereaux et al[22] (permission granted).

pursuit of EBM, and new challenges to practitioners of EBM. Clinicians can provide the best care for their patients by recognizing the need for evidence, appropriately interpreting the quality of evidence, and incorporating patients' values and preferences in the decision-making process.

Paul J. Karanicolas, MD
Department of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, ON, Canada
Regina Kunz, MD, PhD
Basel Institute for Clinical Epidemiology
University Hospital Basel
Basel, Switzerland
Gordon H. Guyatt, MD, MSc, FCCP
Department of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, ON, Canada

The authors benefit in various ways, most nonfinancial, from the successful dissemination of EBM concepts and resources.

Correspondence to: Paul J. Karanicolas, MD, Department of Clinical Epidemiology and Biostatistics, McMaster University Medical Center, 1200 Main St W, Room 2C12, Hamilton, ON, L8N 3Z5, Canada; e-mail: pjkarani@uwo.ca
DOI: 10.1378/chest.08-0068

REFERENCES

1 Antman E, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. JAMA 1992; 268:240–248
2 Guyatt GH, Gutterman D, Baumann MH, et al. Grading strength of recommendations and quality of evidence in clinical guidelines. Chest 2006; 129:174–181
3 Guyatt GH, Sackett DL, Taylor DW, et al. Determining optimal therapy: randomized trials in individual patients. N Engl J Med 1986; 314:889–892
4 Kunz R, Oxman A. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; 317:1185–1190
5 Lacchetti C, Ioannidis J, Guyatt GH. Surprising results of randomized trials: users' guides to the medical literature; a manual of evidence-based clinical practice. Chicago, IL: AMA Press, 2001; 247–269
6 Doroshow J, Locker G, Ifrim I, et al. Prevention of doxorubicin cardiac toxicity in the mouse by N-acetylcysteine. J Clin Invest 1981; 68:1053–1064
7 Unverferth D, Jagadeesh J, Unverferth B, et al. Attempt to prevent doxorubicin-induced acute human myocardial morphologic damage with acetylcysteine. J Natl Cancer Inst 1983; 71:917–920
8 Faden A, Jacobs T, Holaday J. Opiate antagonist improves neurologic recovery after spinal injury. Science 1981; 211: 493–494
9 Bracken M, Shepard M, Collins W, et al. A randomized, controlled trial of methylprednisolone or naloxone in the treatment of acute spinal-cord injury: results of the Second National Acute Spinal Cord Injury Study. N Engl J Med 1990; 322:1405–1411
10 Echt D, Liebson P, Mitchell L, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo: the Cardiac Arrhythmia Suppression Trial. N Engl J Med 1991; 324:781–788
11 Mason J, Peters F. Antiarrhythmic efficacy of encainide in patients with refractory recurrent ventricular tachycardia. Circulation 1981; 63:670–675
12 Rossaint R, Falke K, Lopez F, et al. Inhaled nitric oxide for the adult respiratory distress syndrome. N Engl J Med 1993; 328:399–405
13 Taylor R, Zimmerman J, Dellinger R, et al. Low-dose inhaled nitric oxide in patients with acute lung injury: a randomized controlled trial. JAMA 2004; 291:1603–1609
14 Flaherty K, Kazerooni E, Curtis J, et al. Short-term and long-term outcomes after bilateral lung volume reduction surgery: prediction by quantitative CT. Chest 2001; 119: 1337–1346
15 National Emphysema Treatment Trial Research Group. Patients at high risk of death after lung-volume-reduction surgery. N Engl J Med 2001; 345:1075–1083
16 Rosenqvist M, Brandt J, Schuller H. Long-term pacing in sinus node disease: effects of stimulation mode on cardiovascular morbidity and mortality. Am Heart J 1988; 116:16–22
17 Connolly S, Kerr C, Gent M, et al. Effects of physiologic pacing versus ventricular pacing on the risk of stroke and death due to cardiovascular causes: Canadian Trial of Physiologic Pacing Investigators. N Engl J Med 2000; 342:1385–1391
18 Knox J, Wilmore D, Demling R, et al. Use of growth hormone for postoperative respiratory failure. Am J Surg 1996; 171: 576–580
19 Takala J, Ruokonen E, Webster N, et al. Increased mortality associated with growth hormone treatment in critically ill adults. N Engl J Med 1999; 341:785–792
20 Tonelli M. The limits of evidence-based medicine. Respir Care 2001; 46:1435–1440
21 Guyatt G, Haynes R, Jaeschke R, et al. Users' guides to the medical literature: xxv. Evidence-based medicine: principles for applying the users' guides to patient care. JAMA 2000; 284:1290–1296

22 Devereaux P, Anderson D, Gardner M, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 2001; 323:1218–1222
23 O'Connor A, Legare F, Stacey D. Risk communication in practice: the contribution of decision aids. BMJ 2003; 327: 736–740
24 Mermel LA. Prevention of intravascular catheter-related infections. Ann Intern Med 2000; 132:391–402
25 Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006; 355:2725–2732

Table 2—*Levels of Evidence

Level 1: RCT or metaanalysis
(Lower limit of confidence interval for treatment effect exceeds minimal important benefit)
Level 2: RCT or metaanalysis
(Lower limit of confidence interval for treatment effect overlaps with minimal important benefit)
Level 3: Nonrandomized concurrent cohort study
Level 4: Nonrandomized historic cohort study
Level 5: Case series without control subjects

*Modified with permission from Cook et al.[2]

Counterpoint: Evidence-Based Medicine Lacks a Sound Scientific Base

I firmly believe clinical practice should be based on the best scientific evidence. But how do you define best evidence? Evidence-based medicine (EBM) founders say "identifying the best evidence means using epidemiologic and biostatistical ways of thinking."[1] Table 1 lists five reasons why this approach is scientifically unsound.

Grading

A fundamental premise on which EBM is founded is the ability to grade the quality of research studies. The grading system (levels 1 to 5 evidence) was originally published in a *CHEST* Supplement (Table 2).[2] EBM grading views randomization as not just one important factor but more important than every other component of research methodology. The same concept is rephrased by Sackett et al[3]: "If the study wasn't randomized, we'd suggest that you stop reading it and go on to the next article." EBM grading is based on neither empirical investigation nor rationalist theory. The original article[2] is simply an opinion piece.

Table 1—*Why EBM Lacks a Sound Scientific Base*

EBM grading is detached from scientific theory
(EBM grades homeopathy as level 1 evidence)
Failure of the attempt of logical positivism to demarcate levels of knowledge
(EBM founders do not explain why their system can overcome what proved insurmountable to the foremost epistemologists)
EBM reduces the methodology of science to a single step
(EBM asserts that avoidance of assignment bias cancels every other methodologic error)
EBM confuses statistics for science
(Grading of clinical-practice guidelines is decided by confidence interval and totally ignores breaches of internal validity)
EBM is not internally consistent
(EBM has not tested itself against own standards [an RCT]; thus, by its own standards, EBM is invalid)

There are two reasons why EBM grading is flawed. One, the grading is detached from scientific theory.[2,4] Homeopathy uses drugs in which less than one molecule of active agent is present. Benefit with dilution beyond Avogadro number contradicts pharmacologic theory. A metaanalysis[5] of 89 placebo-controlled trials revealed a combined odds of 2.45 in favor of homeopathy. EBM grades metaanalysis as level 1 evidence but completely ignores scientific theory.[2] There is nothing necessarily wrong with this particular metaanalysis, but the example illustrates how a system that grades findings of all metaanalyses as level 1 evidence[2] is inherently flawed.[6] A grading system that ranks homeopathy as sounder evidence than centuries of pharmacologic science commits the *reductio ad absurdum* fallacy in logic.

Two, attempts at grading of research in other disciplines have failed. The most famous attempt was by the logical positivists.[7] This school contained some of the brightest minds of the early twentieth century. It dominated analytic philosophy of that period. Positivists developed a verifiability criterion, which demarcated "meaningful" from "meaningless" research statements. Popper[8] and others pointed out two fundamental flaws of positivism; thereafter, positivism lost all supporters.[7] EBM retains these two flaws: a dissociation of facts from scientific theory (homeopathy, above), and no empirical testing (see below).

EBM founders have repeatedly revised their grading system.[9] They have, however, never provided reasons why their system is capable of overcoming the problems that proved insurmountable to the logical positivists. Given the defeat of positivism, the leading epistemologists in the world have considered all attempts to grade scientific research as fundamentally flawed.[7,8,10] No field of inquiry, other than clinical medicine, attempts to grade science.

EBM thinking gets even more worrisome. EBM founders say evidence can be "pregraded for validity