Dimensions of Evidence

Alfred O. Berg, MD, MPH

In the January-February issue of the *JABFP*, John Geyman introduced the subject of evidence-based medicine by presenting an overview of the gap between evidence and practice and the challenges ahead as we attempt to fill the gap.[1]

Several parent concerns are driving the evidence-based approach. First, the usual sources of clinical advice, expert opinions, have proved variable and unreliable. When one asks a group of experts a straightforward clinical question, one is as likely as not to get a range of answers rather than a single answer based on evidence. The American Medical Association Diagnosis and Treatment Therapeutic Assessment (DATTA) program in which experts are polled always impresses by the dispersion of the responses as much as by the (uncommon) convergence of opinion. An evidence-based approach should be able to tell us which of the options is best supported by scientific data.

Second, there is tremendous geographic variation in the use of some diagnostic and therapeutic modalities without an evident basis for the variation in the prevalence or incidence of the disease. An evidence-based approach should narrow the variation to those options with the best evidence, regardless of geographic location. Third, there is a gap between practice and evidence. The gap is of at least two types: when we do not have evidence, and when we have evidence but fail to apply it. An evidence-based approach should help make clear which of these problems needs to be addressed.

Fourth, costs of medical care in the United States are higher absolutely and proportionally than in any other industrialized country. An evidence-based approach has potential for constraining costs because standards to accept (and pay for) a new intervention are high. An evidence-based approach could also prove an expensive intervention worthwhile, of course, but there are more examples of the former. Fifth is the growing concern about quality. Is it possible that all processes lead to similar good outcomes? Likely not, and one of the goals of an evidence-based approach is to recognize which strategies would predictably lead to the best outcomes.

This article discusses dimensions of evidence as an historical and philosophical preamble to later articles in the series that will refine concepts of evidence, outcomes, quality, and cost. Using brief clinical scenarios in which one physician asks another a question about a patient, I will explore some of the ways in which physicians define and use evidence in clinical practice. I will use the question "Does it work?" about a clinical intervention, because the "it" can be a fact about etiology or prognosis, a diagnostic test, or a therapeutic or preventive intervention. The central underlying question is whether the advice is "true," forcing us to consider the quality of evidence that supports it. I conclude with a summary of one of the currently popular methods used to answer some of the shortcomings illustrated in the scenarios by outlining the design and construction of clinical practice guidelines.

Illustrative Clinical Scenarios

Trust Me

Dr. Smith: So which statin would you recommend?
Dr. Jones: Atorvastatin is the drug of choice.
(Translation: Does it work? It does if I say so.)

In a busy clinical setting, Dr. Smith asks a straightforward question, and Dr. Jones provides a straightforward answer. Dr. Smith is likely to implement the advice (at least for now) without questioning it. This model dates to the prescientific era when trainees were apprenticed to experienced physicians and simply emulated their practice without questioning why. But this model is also extremely common today, probably the dominant form of teaching in the clinical setting.[2]

The problem from an evidence-based perspective is that the inputs and processes are not ex-

Submitted, revised, 11 February 1998.
From the Department of Family Medicine, University of Washington, Seattle. Address reprint requests to Alfred O. Berg, MD, MPH, Department of Family Medicine, Box 356390, Seattle, WA 98195-6390.

plicit—the questioner has no idea what factors went into making the recommendation and how they were processed before giving the advice. The advice might or might not be based on evidence. The advice might or might not be "true."

Deferral to Authority

Dr. Smith: So which statin would you recommend?
Dr. Jones: I prefer atorvastatin, and the last time I sent someone over to the lipid clinic, that was Dr. Doe's recommendation as well.
(Translation: Does it work? It does if Dr. Expert says so.)

This scenario is a variant of the first, often heard when Dr. Jones is reasonably certain of an answer but wants to bolster her position by referring to an expert's opinion. Expertise and specialized knowledge are highly valued in our society. It is a rare person and a rare physician who questions authorities in specialized areas. The relationship between referring physicians and their consultants is complex, but usually the expert claims to have the answer and the referring physician accepts it. The problem from an evidence-based perspective is similar to that in the first scenario: the inputs and processes are not explicit. The advice might or might not be based on evidence. The advice might or might not be "correct." An evidence-based approach directly confronts the expertise of the expert, in some cases uncomfortably so, by questioning the scientific basis for the recommendation. An evidence-based consultation provides a recommendation and the supporting data rather than the recommendation only.

In My Experience

Dr. Smith: So which statin would you recommend?
Dr. Jones: I don't really like any of them. I have had good experience with diet and cholestyramine and never prescribe HMG-CoA reductase inhibitors.
(Translation: Does it work? It does if it seems to.)

This response is not usually stated quite this way; rather, it comes out as "in my clinical experience." Physicians develop strong opinions about what works and what does not based on their own clinical experience. Thus physicians for many years have used a great many treatments for multiple sclerosis and have convinced themselves that the treatments work based on their own clinical experience. Most such treatments have not withstood the test of a properly conducted clinical trial, yet many are still in use by individual physicians because their experience is limited to a small number of patients, the course of the disease is unpredictable, and most patients seem to improve.

In this scenario we have a physician using as first line a drug that is outmoded. It was probably the drug of choice when the physician left residency, but progress has passed him by. Clinical knowledge tends to decline with time.[3] Year of graduation from residency is an important predictor of the drug formulary used by the average physician.[4]

Clinical experience is the source for many therapies that have proved efficacious, of course, but also has been the source for interventions whose efficacy has been disproved or is unknown. Much of the controversy around alternative medicine can be viewed from this perspective. Because so many human ailments are self-limited, practitioners of aromatherapy, reflexology, and Rolfing are able to persuade themselves that their interventions work when really we have no scientific evidence whether they do or not.

The Pathophysiologic Model

Resident (at 2 AM): This low-risk woman is at term and in labor. Anything else I should be doing right now?
Attending: No, get some rest. The external fetal monitor will collect all the information we need to monitor labor.
(Translation: Does it work? It does if it makes sense that it would.)

This interaction illustrates the major focus of clinical practice and scientific work in the last 150 years. We use many clinical interventions because the basic pathophysiology makes sense, even though we might not have true outcome data to show a positive effect. Electronic fetal monitoring makes all kinds of sense based on what we know about maternal and fetal physiology, the natural history of labor, and how we think physicians make decisions; but the evidence supporting the clinical usefulness of routine electronic fetal monitoring is very thin, with the best quality studies showing the least benefit.[5]

In the past it was considered sufficient to understand the pathophysiologic process behind a condition and prescribe treatments that interrupted the process. We now know that this logical linkage is potentially dangerous, because there are examples of where it has not worked or has caused

harm. The arrhythmia suppression trials are an example—promising drugs were found to suppress arrhythmias, but sometimes at the expense of a patient's life.[6]

Everyone Else is Doing It

Dr. Smith: So I have this 55-year-old man in for a complete physical, and he's asking about screening for prostate cancer.
Dr. Jones (medical director): Make sure you do a digital rectal examination and get a PSA. We'll get sued if you don't and he develops prostate cancer, because screening is the standard in our community.
(Translation: Does it work? It does if everyone agrees that it does.)

This answer is common and potentially dangerous. The problem is that if everyone assumes the answer to be X, but X has never been subjected to properly conducted clinical studies, everyone could be blissfully unaware that they are wrong. We do not know whether the medical director's advice above is right or wrong, because we do not have solid evidence that screening for prostate cancer with prostate-specific antigen is beneficial, although it is very nearly the standard of care in this country. The unquestioned use of radical mastectomy for breast cancer is a sobering example of where the presumption of benefit was so strong for so many decades that the first physicians who questioned it were professionally isolated.

The US Preventive Services Task Force was often surprised to discover the poor quality of evidence supporting preventive interventions already in wide practice. To suggest that we need placebo-controlled randomized trials assessing quality and quantity of life is a radical notion for many clinical conditions. I have learned to be guarded when the consensus on what to do appears strongest.

A Reference Standard?

Dr. Smith: I have a 26-year-old man with his first episode of low back pain. What can I recommend?
Dr. Jones: (silent)
(Translation: Does it work? Well, what should Dr. Jones recommend?)

Much of what we do in clinical medicine has not been subjected to well-designed scientific studies. Randomized controlled trials (RCTs) are the reference standard for interventions, but RCTs are not perfect, and they do not apply to studies of cause, diagnosis, or prognosis. RCTs are not often conducted on patients similar to those encountered in practice, and the selection process and follow-up of patients in an RCT are usually quite different from what you might be able to achieve in a typical practice setting. There are some clinical interventions that will never be subjected to an RCT.

Still, evidence from RCTs is, at least for now, the most scientific point of entry into a discussion with a patient about what his or her options are for a therapeutic or preventive intervention. If no data are available, the discussion has a different end point than if data are available and the main issue is determining whether they apply to the patient in the examining room.

It is incredibly rare, however, that a single RCT, even if well designed, answers the question for all patients in all settings. Medical research almost never provides silver bullets. So if we are concerned that our clinical practice is based on high-quality evidence, we need to ask the questions: Which evidence? What defines high quality? How much do we need?

Clinical Practice Guidelines

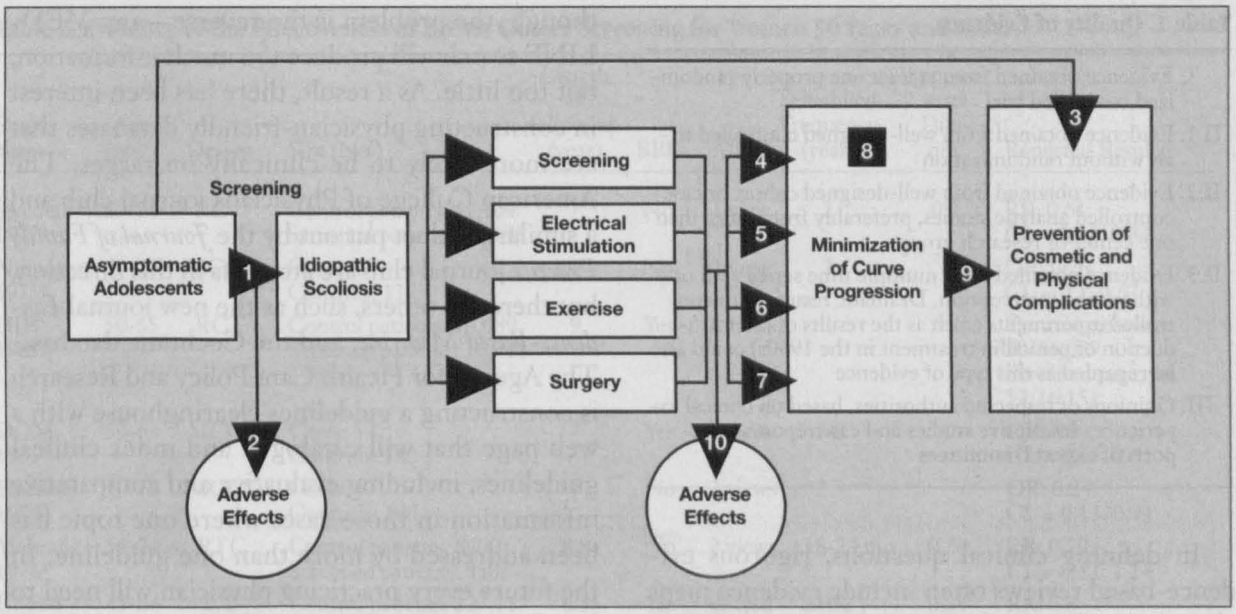
One approach to answering these questions is to develop a systematic way of defining, collecting, analyzing, and summarizing the evidence into a clinical practice guideline. The process of constructing an evidence-based clinical practice guideline is described by many authors, but all such processes have in common several steps:

1. Define the question.
2. Find the evidence.
3. Analyze the evidence.
4. Summarize the evidence.

Define the Question

Defining the question sounds easy, but it can be challenging. A panel of experts constructing a clinical guideline can spend many hours, even days, focusing and refining the question.

David Sackett has proposed the following list of the central tasks of clinical work or where clinical questions come from[7]:
Clinical findings—how to properly gather and interpret findings from the history and physical examination
Etiology—how to identify causes for diseases, including iatrogenesis
Differential diagnosis—how to identify and rank



Linkage: Step in Causal Pathway	Evidence Codes*	Quality of Evidence
1. Accuracy ot screening tests: evidence that physical examination of back can detect curves	II-2	Fair: significant interrater variation, poor reference standard, lack of evidence from physician screening
2. Adverse effects of screening: evidence that screening is associated with an increased risk of complications	III	Poor: most postulated adverse effects have not been evaluated in studies
3. Effectiveness of early detection: evidence that persons detected through screening have better outcomes than those who are not screened	II-3	Poor: uncontrolled studies based on time trends after initiation of screening, failure to control for confounding temporal factors
4. Braces	II-2, II-3	Poor: selection bias, lack of internal control groups
5. Lateral electrical surface stimulation	II-2, II-3	(most studies), inadequate follow-up,
6. Exercise	I, II-3	small sample sizes, lack of health outcome measures
7. Surgery	II-2, II-3	
8. Curve progression: evidence that curves detected on screening are destined to progress to curves of clinical significance	II-3	Fair: significant number of patients unavailable for follow-up, variable measures of progression
9. Complications of curve progression: evidence that persons with scoliosis are more likely to experience back complaints, psychosocial effects, disability	II-3	Poor: studies generally lack control groups, have high attrition rates, include mixture of patients with different problems, and use variable measures to judge outcome
10. Adverse effects of treatment: evidence that treatment is associated with an increased risk of complications	III	Poor: most postulated adverse effects have not been evaluated in studies

Figure 1. Causal pathway for scoliosis screening. From the US Preventive Services Task Force.[8]

*See Table 1 for explanation of evidence codes.

the alternatives by likelihood, seriousness, and treatability

Diagnostic tests—how to select and interpret diagnostic tests, considering their precision, accuracy, acceptability, cost, and safety

Prognosis—how to estimate the patient's likely course with time and anticipate likely complications

Therapy—how to select treatments to offer patients that do more good than harm and that are worth the efforts and costs of using them

Prevention—how to reduce the chance of disease by identifying and modifying risk factors and how to diagnose early disease by screening

For each of these clinical tasks, there are four elements of a well-formulated question:

1. What is the patient or problem being addressed?
2. What is the intervention?
3. What are the alternatives?
4. What are the outcomes?

For example, if you were looking at a therapy for heart failure, the intervention might be an angiotensin-converting enzyme inhibitor; an alternative, diuretics alone; and the outcomes, the correction of the heart failure or mortality.

Table 1. Quality of Evidence.

I.	Evidence obtained from at least one properly randomized controlled trial
II.1.	Evidence obtained from well-designed controlled trials without randomization
II.2.	Evidence obtained from well-designed cohort or case-controlled analytic studies, preferably from more than one center or research group
II.3.	Evidence obtained from multiple time series with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of the introduction of penicillin treatment in the 1940s) could also be regarded as this type of evidence
III.	Opinions of respected authorities, based on clinical experience, descriptive studies and case reports, or reports of expert committees

In defining clinical questions, rigorous evidence-based reviews often include evidence maps or causal pathways, laying out in advance how a clinical problem is thought to work. This map then directs the literature review. Figure 1 presents a causal pathway for a secondary prevention intervention—screening for adolescent idiopathic scoliosis—used by the US Preventive Services Task Force in conducting an evidence-based review of whether screening worked.[8] In the causal pathway each numbered linkage defines a clinical question for which an evidence search was conducted (the results of the search also appear in the figure).

Find the Evidence

Once the clinical questions are clear, one is ready to gather and summarize the evidence. The important message here is that the methods need to be systematic, thorough, and explicit, so that there is some accountability to the process. MEDLINE is the best-known database, but it is important to point out that MEDLINE targets a minority of the biomedical journals worldwide (although arguably it indexes nearly all the important ones), and that not all MEDLINE-based search products work in the same way. There are more than 30 proprietary versions available, and the same search strategy will yield slightly different results on each. Further, some relevant evidence might be unpublished, it might have been indexed incorrectly so that a search does not capture it, or it even might have been published before 1966 and predate MEDLINE.

For these reasons, really comprehensive searches use databases and strategies that go beyond MEDLINE. For the individual physician, though, the problem is the reverse—any MEDLINE search will produce too much information, not too little. As a result, there has been interest in constructing physician-friendly databases that are more likely to be clinically on target. The American College of Physicians journal club and a similar product put out by the *Journal of Family Practice* journal club are attempts in this direction, but there are others, such as the new journal *Evidence-Based Medicine*, and the Cochrane database. The Agency for Health Care Policy and Research is constructing a guidelines clearinghouse with a web page that will catalogue and index clinical guidelines, including evaluative and comparative information in those cases where one topic has been addressed by more than one guideline. In the future every practicing physician will need to be adept at using electronic databases and determining which databases fit which clinical needs best.

Analyze the Evidence

Once assembled, the evidence needs to be reviewed for applicability and quality. Reviewing the quality of evidence has been the subject of countless journal clubs and seminars for many years. This review process is where everything physicians learn about evaluating published research comes into practice. The standards for judging the quality of research on diagnostic tests, prognosis, treatment, and adverse events are very high, indeed, and getting higher every year. As these standards are applied, fewer articles pass muster. Those working in evidence-based medicine now are nearly unanimous in demanding high-quality randomized trials for interventions. Unfortunately, case studies, case series, observational studies, and uncontrolled trials still dominate in the medical literature on interventions.

For the foreseeable future, high-quality evidence will not be available for many important clinical questions. In these circumstances an evidence-based approach at minimum requires explicit criteria for categorizing the design (eg, case-control study, randomized controlled trial) and judging the quality of a study. Many resources are available to help make the judgments. The grading of evidence used by the US Preventive Services Task Force is a typical example of categorizing studies (Table 1).[9] Category-specific criteria for judging the quality of a study are also available

Table 2. Evidence of the Effectiveness of Breast Cancer Screening for Women 50 Years and Older.*

Study, Year	Age (year)	Design	Size (No.)	Length of follow-up (year)	BPE	MGY	Scheduled Frequency (year)	Approximate Dilution	Reported Results
Sweden, 1977†	50-74	RCT	Control patients: 41,104 Screened patients: 58,148	7	No	1 view	2	0.2	RR: 0.61 CL = 0.44, 0.84
Nijmegen, 1975‡	50-64	MCCS	Cases: 27 Controls: 135	7	No	1 view	2		OR: 0.26 CL = 0.1, 0.67
HIP, 1963§	50-65	RCT	Control patients: 16,089 Screened patients: 16,151	9	Yes	2 views	1	0.4	Control patients: 80/16,089 Screened patients: 52/16,151
DOM, 1975‖	50-64	MCCS	Cases: 54 Controls: 162	7	Yes	2 views	1-2		OR: 0.31 CL = 0.15, 0.65
Florence, 1979¶	40-70	MCCS	Cases: 57 Controls: 186	7	No	2 views	2.5		OR: 0.24 CL = 0.13, 0.42
Malmok 1976#	55-74	RTC	Control patients: 8490 Screened patients: 8507	8.8	No	2 views	18-24 mo	0.28	RR: 0.79 CL = 0.51, 1.24
United Kingdom, 1979**	45-64	CT	Control patients: 127,117 Screened patients: 45,841	6.5	Yes	Mixed	BPE: 1 MGY: 2	0.34	RR: 0.8 CL = 0.64, 1.01

Adapted from Eddy DM. A manual for assessing health practices and designing practice policies: the explicit approach. Philadelphia: American College of Physicians, 1992:45. With permission of the American College of Physicians.

BPE - breast physician examination, MGY - mammography, RCT - randomized controlled trial, RR - relative risk, CL - confidence limits, MCCS - matched case-control study, OR - odds ratio, CT - controlled trial.

*To permit comparisons, results are reported for 6 to 10 years of follow-up. Longer follow-up results are available from some studies. Reported results incorporate any adjustments performed by the investigator.

†Tabar L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. Br J Cancer 1987;55:547-51.

‡Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. Lancet 1984;1:1222-4.

Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F. Mammographic screening and breast cancer mortality: age-specific effects in Nijmegen Project, 1975-82 [Letter]. Lancet 1985:865-6.

§Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. J Natl Cancer Inst 1982;69:349-55.

Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. Natl Cancer Inst Monogr 1985;67: 65-74.

Shapiro S, Venet W, Strax P. Periodic screening for breast cancer: the Health Insurance Plan project and its sequelae 1963-1986. Baltimore: Johns Hopkins University Press, 1988.

‖Collette HJ, Day NE, Romback JJ, deWaard F. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. Lancet 1984;1:1224-6.

¶Palli D, Del Turco MR, Buiatti E, Carli S, Ciatto S, Toscani L, et al. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). Int J Cancer 1986;38:501-4.

#Andersson I, Aspergren K, Janson L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BM J 1988;297:943-8.

**First results on mortality reduction in the UK trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Lancet 1988;2:411-6.

from the many sources that address critical reading skills.[7]

Summarize the Evidence

The last step in conducting an evidence review is to summarize what you have collected in some useful way. There are two products to discuss here: evidence tables and outcomes tables (also referred to as balance sheets).

An evidence table is simply a systematic way of displaying information from multiple studies so that they can be compared. For example, the table would have a row for each study included in the database, with the columns containing study characteristics, such as patient population studied, duration, specific interventions, outcomes, and actual data for groups and subgroups. Evidence tables allow a user to scan quickly those study characteristics that are relevant to the question at hand. A sample evidence table summarizing some of the principal breast cancer screening trials is presented in Table 2.[10]

An outcomes table takes the materials from the evidence tables and summarizes them even more concisely. How to do this using meta-analysis and other techniques will be the subject of future articles in this series, but in general, an outcomes table displays the alternative interventions and out-

Table 3. A Balance Sheet for Outcomes of 10 Years of Annual Breast Cancer Screening with Breast Physical Examination and Mammography in Women 55 to 65 Years Old.

	Event	Probability or Magnitude Without Screening	Probability or Magnitude With Screening	Differences Caused by Screening
Background	Develop breast cancer in 10-y period (probability)	2.33% (233/10,000)	2.33% (233/10,000)	0% (0/10,000)
Benefits	Die (ever) from a breast cancer that develops in 10-y period (probability)	1.23% (123/10,000)	0.735 (95% CL - 0.41%, 0.93%) (73/10,000)	0.5% (95% CL - 0.3%, 0.82%) (50/10,000)
	Reassurance* from knowledge that probability of cancer is decreased by 1.7% (probability)	0 (0/10,000)	78.61%† (7861/10,000)	78.61% (7861/10,000)
Harms	Number of physical and mammographic examinations (inconvenience, anxiety, discomfort)	0	10	10
	False-positive result during 10 y (probability)	0 (0/10,000)	20% (2000/10,000)	20% (2000/10,000)
	New breast cancer caused by 10 y of radiation (probability)	0 (0/10,000)	0.0004% (1/250,000)	0.0004% (1/250,000)

From Eddy DM. A manual for assessing health practices and designing practice policies: the explicit approach. Philadelphia: American College of Physicians, 1992: 56. With permission of the American College of Physicians.

* Without screening, the probability that a woman will have breast cancer during the 10 years is about 2.33%. If the screening test results are negative, the probability that she will have breast cancer during the 10 years is decreased to about 0.6%, Thus, if a woman has 10 examinations with negative results, her probability of developing breast cancer in the 10 years is decreased by 1.73% (2.33% - 0.6% = 1.73%). The probability that all 10 tests will have negative results is 78.61%.

†The probability that all 10 tests results will be negative.

comes in a way that the end users can understand. A classic outcomes table covering breast cancer screening in women aged 55 to 65 years is presented in Table 3.[10] The natural history and benefits and harms from the interventions are the rows, and the alternative interventions are the columns. A complete outcomes table will present harms as well as benefits, using data collected and summarized with the same care and rigor for both.

It is important to point out that the numbers in an outcomes table are not directive; that is, reasonable physicians and patients might make different decisions based on the individual values they place on the risks and outcomes. The patient's informed preferences are extremely important, a subject to be addressed in a future article in this series.

Conclusion

Practicing evidence-based medicine requires a shift of the clinical thought processes that most physicians were trained to use. In the past, clinical decisions and clinical advice have relied on clinical experience, expert opinions, collegial relationships, pathophysiology, common sense, community standards, published materials, and other sources. The practice of evidence-based medicine uses the same sources for clinical advice but passes all of them through the filter of the following question: "On what evidence is the advice based?" A properly constructed clinical practice guideline has the potential to serve as a lens for the evidence that does exist (ie, that has passed through the filtering question), focusing it on specific clinical issues in an explicit and accountable way.

Not all evidence is of the same quality. We know that evidence from a properly conducted randomized clinical trial is more likely to be true than evidence based on one random physician's clinical experience or personal opinion; but in giving and receiving advice, we rarely pause to consider what the quality of the underlying evidence might be. I recommend pausing more often.

References

1. Geyman JP. Evidence-based medicine in primary care: an overview. J Am Board Fam Pract 1998;11: 46-56.
2. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? Ann Intern Med 1985;103:596-9.
3. Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Norcini JJ, et al. Changes over time in the knowledge base of practicing internists. JAMA 1991;266:1103-7.
4. Sackett DL, Haynes RB, Taylor DW, Gibson ES, Roberts RS, Johnson AL. Clinical determinants of

the decision to treat primary hypertension. Clin Res 1977;24:648.
5. Vintzileos AM, Nochimson DJ, Guzman ER, Knuppel RA, Lake M, Schifrin BS. Intrapartum electronic fetal heart rate monitoring versus intermittent auscultation: a meta-analysis. Obstet Gynecol1995;85:149-55.
6. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. New Engl J Med 1991,324:781-8.
7. Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone, 1997.
8. Screening for adolescent idiopathic scoliosis. Policy statement. US Preventive Services Task Force JAMA 1993;269:2664-6.
9. US Preventive Services Task Force. Guide to clinical preventive services: report of the US Preventive Services Task Force. 2nd edition. Baltimore: Williams & Wilkins, 1996.
10. Eddy DM. A manual for assessing health practices & designing practice policies: the explicit approach. Philadelphia: American College of Physicians, 1992.