The Medical Literature ▰▰▰▰▰▰▰

Users' Guides to the Medical Literature

II. How to Use an Article About Therapy or Prevention

B. What Were the Results and Will They Help Me in Caring for My Patients?

Gordon H. Guyatt, MD, MSc; David L. Sackett, MD, MSc; Deborah J. Cook, MD, MSc;
for the Evidence-Based Medicine Working Group

CLINICAL SCENARIO

You are a general internist who is asked to see a 65-year-old man with controlled hypertension and a 6-month history of atrial fibrillation resistant to cardioversion. Although he has no evidence for valvular or coronary heart disease, the family physician who referred him to you wants your advice on whether the benefits of long-term anticoagulants (to reduce the risk of embolic stroke) outweigh their risks (of hemorrhage from anticoagulant therapy). The patient shares these concerns and doesn't want to receive a treatment that would do more harm than good. You know that there have been randomized trials of warfarin for nonvalvular atrial fibrillation and decide that you'd better review one of them.

THE SEARCH

The ideal article addressing this clinical problem would include patients with nonvalvular atrial fibrillation and would compare the effect of warfarin and a control treatment, ideally a placebo, on the risk of emboli (including embolic stroke) and also on the risk of the complications of anticoagulation. Randomized, double-blind studies would provide the strongest evidence.

In the software program GRATEFUL MED you select a Medical Subject Heading (MeSH) that identifies your population, "atrial fibrillation," another that specifies the intervention, "warfarin," and a third that specifies the outcome of interest, "stroke" (which the software automatically converts to "explode cerebrovascular disorders" meaning that all articles indexed under cerebrovascular disorders or its subheadings are potential targets of the search), while restricting the search to English-language studies. To ensure that, at least on your first pass, you identify only the highest quality studies, you include the methodological term "randomized controlled trial (PT)" (PT stands for publication type). The search yields nine articles. Three are editorials or commentaries, one addresses prognosis, and one focuses on quality of life for patients receiving anticoagulants. You decide to read the most recent of the four randomized trials.[1]

Reading the study, you find it meets the validity criteria you learned about in a prior article in this series.[2] To answer your patient's and the referring physician's concerns, however, you need to delve further into the relation between benefits and risks.

INTRODUCTION

The previous article in this series dealt with whether a study of effectiveness of therapy was valid (Table 1). In this installment, we will show you how to proceed further to understand and use the results of valid studies of therapeutic interventions. We have summarized calculations in the Tables for easy reference.

What Were the Results?

How Large Was the Treatment Effect?—Most frequently, randomized clinical trials carefully monitor how often patients experience some adverse event or outcome. Examples of these dichotomous outcomes (yes or no outcomes that either happen or don't happen) include cancer recurrence, myocardial infarction, and death. Patients either do or do not suffer an event, and the article reports the proportion of patients who develop such events. Consider, for example, a study in which 20% (0.20) of a control group died, but only 15% (0.15) of those receiving a new treatment died. How might these results be expressed? Table 2 provides a summary of ways of presenting the effects of therapy.

One way would be as the absolute difference (known as the *absolute risk reduction* or risk difference), between the proportion who died in the control group (X) and the proportion who died in the treatment group (Y), or $X - Y =$

0.20−0.15=0.05. Another way to express the impact of treatment would be as a *relative risk* (RR): the risk of events among patients receiving the new treatment, relative to that among controls, or Y/X =0.15/0.20=0.75.

The most commonly reported measure of dichotomous treatment effects is the complement of this RR, and is called the *relative risk reduction* (RRR). It is expressed as a percent: $[1−(Y/X)] \times 100\% = [1−0.75] \times 100\%=25\%$. An RRR of 25% means that the new treatment reduced the risk of death by 25% relative to that occurring among control patients; the greater the RRR, the more effective the therapy.

How Precise Was the Estimate of Treatment Effect?—The true risk reduction can never be known; all we have is the estimate provided by rigorous controlled trials, and the best estimate of the true treatment effect is that observed in the trial. This estimate is called a "point estimate" in order to remind us that although the true value lies somewhere in its neighborhood, it is unlikely to be precisely correct. Investigators tell us the neighborhood within which the true effect likely lies by the statistical strategy of calculating confidence intervals (CIs).[3]

We usually (though arbitrarily) use the 95% CI, which can be simply interpreted as defining the range that includes the true RRR 95% of the time. You'll seldom find the true RRR toward the extremes of this interval, and you'll find the true RRR beyond these extremes only 5% of the time, a property

of the CI that relates closely to the conventional level of "statistical significance" of $P<.05$. We illustrate the use of CIs in the following examples.

If a trial randomized 100 patients each to treatment and control groups, and there were 20 deaths in the control group and 15 deaths in the treatment group, the authors would calculate a point estimate for the RRR of 25%: X=20/100 or 0.20, Y=15/100 or 0.15, and $[1−(Y/X)] \times 100\%=[1−0.75] \times 100=25\%$. You might guess, however, that the true RRR might be much smaller or much greater than this 25%, based on a difference of just five deaths. In fact, you surmise that the treatment might provide no benefit (an RRR of 0%) or even harm (a negative RRR). And you would be right—in fact, these results are consistent with both an RRR of −38% (that is, patients given the new treatment might be 38% more likely to die than control patients), and an RRR of nearly 59% (that is, patients subsequently receiving the new treatment might have a risk of dying almost 60% less than that of the risk in those who are not treated). In other words, the 95% CI on this RRR is −38% to 59%, and the trial really hasn't helped us decide whether to offer the new treatment. What sort of study would be more helpful?

What if the trial enrolled not 100 patients per group, but 1000 patients per group, and observed the same event rates as before, so that there were 200 deaths in the control group (X=200/1000=0.20) and 150 deaths in the treatment group (Y=150/1000=0.15). Again, the point estimate of the RRR is 25%: $[1−(Y/X)] \times 100\%=[1−(0.15/0.20)] \times 100\%=25\%$. In this larger trial, you might think that the true reduction in risk is much closer to 25% and, again, you would be right; the 95% CI on the RRR for this set of results is all on the positive side of 0 and runs from 9% to 41%.

What these examples show is that the larger the sample size of a trial, the larger the number of outcome events and the greater our confidence that the true RRR (or any other measure of efficacy) is close to what we have observed. In the second example above, the lowest plausible value for the RRR was 9% and the highest value 41%. The point estimate—in this case 25%—is the one value most likely to represent the true RRR. As one considers values farther and farther from the point estimate, they become less and less

consistent with the observed RRR. By the time one crosses the upper or lower boundaries of the 95% CI, the values are extremely unlikely to represent the true RRR, given the point estimate (that is, the observed RRR).

The Figure represents the CIs around the point estimate of an RRR of 25% in these two examples, with a risk reduction of 0 representing no treatment effect. In both scenarios the point estimate of the RRR is 25%, but the CI is far narrower in the second scenario.

It is evident that the larger the sample size, the narrower the CI. When is the sample size big enough?[4] In a "positive" study—a study in which the authors conclude that the treatment is effective—one can look at the lower boundary of the CI. In the second example, this lower boundary was +9%. If this risk reduction (the lowest that is consistent with the study results) is still important, or "clinically significant," (that is, it is large enough for you to want to offer it to your patient), then the investigators have enrolled sufficient patients. If, on the other hand, you do not consider an RRR of 9% clinically significant, then the study cannot be considered definitive, even if its results are statistically significant (that is, they exclude a risk reduction of 0). Keep in mind that the probability of the true value being less than the lower boundary of the CI is only 2.5%, and that a different criterion for the CI (a 90% CI, for instance) might be as or more appropriate.

The CI also helps us interpret "negative" studies in which the authors have concluded that the experimental treatment is no better than control therapy. All we need do is look at the upper boundary of the CI. If the RRR at this upper boundary would, if true, be clinically important, the study has failed to exclude an important treatment effect. In the first example we presented in this section, the upper boundary of the CI was an RRR of 59%. Clearly, if this represented the truth, the benefit of the treatment would be substantial, and we would conclude that although the investigators had failed to prove that experimental treatment was better than placebo, they also had failed to prove that it was not; they could not exclude a large, positive treatment effect. Once again the clinician must bear in mind the proviso about the arbitrariness of the choice of 95% boundaries for the CI. A reason-

Table 1.—Readers' Guides for an Article About Therapy

Are the results of the study valid?
Primary guides:
   Was the assignment of patients to treatments randomized?
   Were all patients who entered the trial properly accounted for and attributed at its conclusion?
      Was follow-up complete?
      Were patients analyzed in the groups to which they were randomized?
Secondary guides:
   Were patients, health workers, and study personnel "blind" to treatment?
   Were the groups similar at the start of the trial?
   Aside from the experimental intervention, were the groups treated equally?
**What were the results?**
   How large was the treatment effect?
   How precise was the estimate of the treatment effect?
Will the results help me in caring for my patients?
   Can the results be applied to my patient care?
   Were all clinically important outcomes considered?
   Are the likely treatment benefits worth the potential harms and costs?

Table 2.—Introducing Some Measures of the Effects of Therapy

Risk without therapy (baseline risk): X	Risk with therapy: Y	Absolute risk reduction (risk difference): X−Y	Relative risk: Y/X	Relative risk reduction (RRR): [1−(Y/X)]×100% or [(X−Y)/X]×100%	95% confidence interval for the RRR:
20/100=0.20 or 20%	15/100=0.15 or 15%	0.20−0.15=0.05	0.15/0.20=0.75	[1−0.75]×100%=25% [0.05/0.20]×100%=25%	−38% to +59%
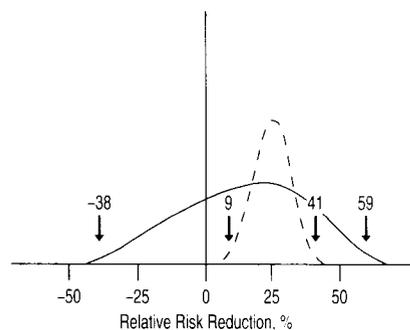
The solid line represents the confidence interval around the first example in which there were 100 patients per group and the number of events in the active and control groups were two and four, respectively. The broken line represents the confidence interval around the second example in which there were 1000 patients per group and the number of events in the active and control groups were 20 and 40, respectively.

able alternative, a 90% CI, would be somewhat narrower.

What can the clinician do if the CI around the RRR is not reported in the article? There are three approaches, and we present them in order of increasing complexity. The easiest approach is to examine the P value. If the P value is exactly .05, then the lower bound of the 95% confidence limit for the RRR has to lie exactly at 0 (an RR of 1), and you cannot exclude the possibility that the treatment has no effect. As the P value decreases below .05, the lower bound of the 95% confidence limit for the RRR rises above 0.

A second approach, involving some quick mental arithmetic or a pencil and paper, can be used when the article includes the value for the standard error (SE) of the RRR (or of the RR). This is because the upper and lower boundaries of the 95% CI for an RRR are the point estimate plus and minus twice this SE.

The third approach involves calculating the CIs yourself[5] or asking the help of someone else (a statistician, for instance) to do so. Once you obtain the CIs, you know how high and low the RRR might be (that is, you know the precision of the estimate of the treatment effect) and can interpret the results as described above.

Not all randomized trials have dichotomous outcomes, nor should they. For example, a new treatment for patients with chronic lung disease may focus on increasing their exercise capacity. Thus, in a study of respiratory muscle training for patients with chronic airflow limitation, one primary outcome measured how far patients could walk in 6 minutes in an enclosed corridor.[6] This 6-minute walk improved from an average of 406 to 416 meters (up 10 meters) in the experi-

mental group receiving respiratory muscle training, and from 409 to 429 (up 20 meters) in the control group. The point estimate for improvement in the 6-minute walk due to respiratory muscle training therefore was negative, at -10 meters (or a 10-meter difference in favor of the control group).

Here too you should look for the 95% CIs around this difference in changes in exercise capacity and consider their implications. The investigators tell us that the lower boundary of the 95% CI was -26 meters (that is, the results are consistent with a difference of 26 meters in favor of the control treatment) and the upper boundary was $+5$ meters. Even in the best of circumstances, adding 5 meters to the 400 recorded at the start of the trial would not be important to the patient, and this result effectively excludes a clinically significant benefit of respiratory muscle training as applied in this study.

Having determined the magnitude and precision of the treatment effect, readers now can turn to the final question of how to apply the article's results to their patients and clinical practice.

Will the Results Help Me in Caring for My Patients?

Can the Results Be Applied to My Patient Care?—The first issue to address is how confident you are that you can apply the results to a particular patient or patients in your practice. If the patient would have been enrolled in the study had she been there—that is, she meets all the inclusion criteria, and doesn't violate any of the exclusion criteria—there is little question that the results are applicable. If this is not the case, and she would not have been eligible for the study, judgment is required. The study result probably applies even if, for example, she was 2 years too old for the study, had more severe disease, had previously been treated with a competing therapy, or had a comorbid condition. A better approach than rigidly applying the study's inclusion and exclusion criteria is to ask whether there is some compelling reason why the results should *not* be applied to the patient. A compelling reason usually won't be found, and most often you can generalize the results to your patient with confidence.

A final issue arises when our patient fits the features of a subgroup of patients in the trial report. In articles reporting the results of a trial (especially when the treatment doesn't appear to be efficacious for the average patient), the authors may have examined a large number of subgroups of patients at different stages of their illness, with dif-

ferent comorbid conditions, with different ages at entry, and the like. Quite often these subgroup analyses were not planned ahead of time, and the data are simply "dredged" to see what might turn up. Investigators may sometimes overinterpret these "data-dependent" analyses as demonstrating that the treatment really has a different effect in a subgroup of patients—those who are older or sicker, for instance, may be held up as benefitting substantially more or less than other subgroups of patients in the trial. You can find guides for deciding whether to believe these subgroup analyses,[7] summarized as follows: the treatment is really likely to benefit the subgroup more or less than the other patients if the difference in the effects of treatment in the subgroups (1) is large; (2) is very unlikely to occur by chance; (3) results from a analysis specified as a hypothesis before the study began; (4) was one of only a very few subgroup analyses that were carried out; and (5) is replicated in other studies. To the extent that the subgroup analysis fails these criteria, clinicians should be increasingly skeptical about applying them to their patients.

Were All Clinically Important Outcomes Considered?—Treatments are indicated when they provide important benefits. Demonstrating that a bronchodilator produces small increments in forced expired volume in patients with chronic airflow limitation, that a vasodilator improves cardiac output in heart failure patients, or that a lipid-lowering agent improves lipid profiles does not necessarily provide a sufficient reason for administering these drugs. What is required is evidence that the treatments improve outcomes that are important to patients, such as reducing shortness of breath during the activities required for daily living, avoiding hospitalization for heart failure, or decreasing the risk of myocardial infarction. We can consider forced expired volume in 1 second, cardiac output, and the lipid profile "substitute end points." That is, the authors have substituted these physiologic measures for the important outcomes (shortness of breath, hospitalization, or myocardial infarction), usually because to confirm benefit on the latter they would have had to enroll many more patients and followed them for far longer periods of time.

A dramatic recent example of the danger of substitute end points was found in the evaluation of the usefulness of antiarrhythmic drugs following myocardial infarction. Because such drugs had been shown to reduce abnormal ventricular depolarizations (the substitute end points) in the short run, it made

Table 3.—Two Men With Contrasting Prognoses Following Myocardial Infarction

If the risk of death at 1 year without therapy (baseline risk) is: X	And the relative risk of death with therapy (a β blocker) is: Y/X	And the relative risk reduction is: $[1-(Y/X)]\times100\%$ or $[(X-Y)/X]\times100\%$	Then the risk of death with treatment is: Y	And the absolute risk reduction is: X−Y	And the number needed to be treated to prevent one event is: 1/(X−Y)
1% or 0.01	75% or 0.75	25%	0.01×0.75=0.0075	0.01−0.0075=0.0025	1/0.0025=400
10% or 0.10	75% or 0.75	25%	0.10×0.75=0.075	0.10−0.075=0.025	1/0.025=40

Table 4.—Incorporating Side Effects Into the Number Needed to Be Treated

If the risk of death at 1 year without therapy (baseline risk) is: X	And the risk of death with propranolol is: Y	Then the absolute risk reduction is: X−Y	And the number needed to be treated to prevent one event is: 1/(X−Y)	And if the incidence of clinically important fatigue on propranolol is:	Then the number of fatigued patients per life saved is:
1% or 0.01	0.01×0.75=0.0075	0.01−0.0075=0.0025	1/0.0025=400	10% or 0.10	400×0.1=40
10% or 0.10	0.10×0.75=0.075	0.10−0.075=0.025	1/0.025=40		40×0.1=4

sense that they should reduce the occurrence of life-threatening arrhythmias in the long run. A group of investigators performed randomized trials on three agents (encainide, flecainide, and moricizine) previously shown to be effective in suppressing the substitute end point of abnormal ventricular depolarizations in order to determine whether they reduced mortality in patients with asymptomatic or mildly symptomatic arrhythmias following myocardial infarction. The investigators had to stop the trials when they discovered that mortality was substantially higher in patients receiving antiarrhythmic treatment than in those receiving a placebo.[8,9] Clinicians relying on the substitute end point of arrhythmia suppression would have continued to administer the three drugs, to the considerable detriment of their patients.

Even when investigators report favorable effects of treatment on one clinically important outcome, clinicians must take care that there are no deleterious effects on other outcomes. For instance, as this series was in preparation, the controversy continued over whether reducing lipids unexpectedly increases noncardiovascular causes of death.[10] Cancer chemotherapy may lengthen life but may also decrease its quality. Finally, surgical trials often document prolonged life for those who survive the operation (yielding higher 3-year survival in those receiving surgery), but an immediate risk of dying during or shortly after surgery. Accordingly, users of the reports of surgical trials should look for information on immediate and early mortality (typically higher in the surgical group) in addition to longer-term results.

Are the Likely Treatment Benefits Worth the Potential Harm and Costs?—If the article's results are generalizable to your patient and its outcomes are important, the next question concerns whether the probable treatment benefits are worth the effort that you and your patient must put into the enterprise. A 25% reduction in the risk of death may sound quite impressive, but its impact on your patient and practice may nevertheless be minimal. This notion is illustrated using a concept called "number needed to treat" (NNT).[11]

The impact of a treatment is related not only to its RRR, but also to the risk of the adverse outcome it is designed to prevent. β-Blockers reduce the risk of death following myocardial infarction by approximately 25%, and this RRR is consistent across subgroups, including those at higher and lower "baseline" risk of recurrence and death when they are untreated. Table 3 considers two patients with recent myocardial infarctions.

First, consider a 40-year-old man with a small infarct, normal exercise capacity, and no sign of ventricular arrhythmia who is willing to stop smoking, begin exercising, lose weight, and take aspirin daily. This individual's risk of death in the first year after infarction may be as low as 1%. β-Blockers would reduce this risk by a quarter, to 0.75%, for an absolute risk reduction of 0.25% or 0.0025. The inverse of this absolute risk reduction (that is, 1 divided by the absolute risk reduction) equals the number of such patients we'd have to treat in order to prevent one event (in this case, to prevent one death following a mild heart attack in a low-risk patient). In this case, we would have to treat 400 such patients for 1 year to save a single life (1/0.0025=400).

An older man with limited exercise capacity and frequent ventricular extrasystoles who continues to smoke following his infarction may have a risk of dying in that next year as high as 10%. A 25% risk reduction for death in such a high-risk patient generates an absolute risk reduction of 2.5% or 0.025, and we would have to treat only 40 such individuals for 1 year to save a life (1/0.025=40).

These examples underscore a key element of the decision to start therapy: before deciding on treatment, we must consider our patient's risk of the adverse event if left untreated. For any given RRR, the higher the probability that a patient will experience an adverse outcome if we don't treat, the more likely the patient will benefit from treatment, and the fewer such patients we need to treat to prevent one event. Thus, both patients and our own clinical efficiency benefit when the NNT to prevent an event is low.

We might not hesitate to treat even as many as 400 patients to save one life if the treatment were cheap, easy to apply and comply with, and safe. In reality, however, treatments usually are expensive and they carry risks. When these risks or adverse outcomes are documented in trial reports, users can apply the NNT to judge both the relative benefits and costs of therapy. If, for instance, β-blockers cause clinically important fatigue in 10% of the patients who use them, the NNT to cause fatigue is 1/0.10 or 10. This is shown in Table 4, where it is seen that a policy of treating low-risk patients after myocardial infarction (NNT=400 to prevent one death) will result in 40 being fatigued for every life saved. On the other hand, a policy of treating just high-risk patients will result in four being fatigued for every life saved.

Clinicians don't, however, treat groups of patients uniformly. Rather, we consider individual responses and tailor our therapy accordingly. One response to the problem of common, relatively minor side effects (such as fatigue) is to discontinue therapy in patients suffering from that problem. If we think of fatigued low-risk patients as a group, we would make 400 patients fatigued to save a life, a trade-off that probably wouldn't be worth it. By discontinuing treatment in these people, we can treat the remainder without making anyone fatigued.

We cannot apply this approach, however, to severe, episodic events. Examples include the risk of bleeding in patients given anticoagulants, throm-

Table 5.—Summary of the Effect of Warfarin Therapy on Patients With Nonvalvular Atrial Fibrillation*

If the risk of stroke at 1 year without therapy (baseline risk) is: X	And the risk of stroke with treatment is: Y	Then the absolute risk reduction is: X−Y	And the number needed to be treated to prevent one stroke is: 1/(X−Y) (95% confidence interval)	And if the incidence of clinically important bleeding on warfarin is:	Then the number of bleeds per stroke prevented is:
0.043	0.009	0.043−0.009=0.034	1/0.034=30 (26 to 45)	0.01	29×0.01=0.29

*Data from Ezekowitz et al.[1]

bolytic agents, or aspirin, or the risk of rare but devastating drug reactions. In each of these examples the number of adverse events per life saved (or, if the events are rare enough, the number of lives saved per adverse event) can provide a compelling picture of the trade-offs associated with the intervention.

RESOLUTION OF THE SCENARIO

In the randomized trial of warfarin in nonvalvular atrial fibrillation that you selected for reading (Ezekowitz et al[1]), 260 patients received warfarin and 265 received placebo. The results are summarized in Table 5.

Over the next 1½ years, just four of the former (0.9% per year), but 19 of the latter (4.3% per year) suffered cerebral infarction. Thus, the RRR is (0.043−0.009)/0.043=79%, the absolute risk reduction is 0.043−0.009= 0.034, and the NNT to prevent one stroke is 1/0.034 =29 (or approximately 30). Applying CIs to this NNT, the NNT could be (using the lower boundary of the CI around the RRR, which was 0.52) as great as 45, or (using the upper boundary of the CI around the RRR, which was 0.90) as few as 26. Now, you know that warfarin is a potentially dangerous drug, and that about 1% of patients on this treatment will suffer clinically important bleeding as a result of treatment each year.[12] Therefore, there will be one episode of bleeding in every 100 treated patients, and if the NNT to prevent a stroke is 30, then for every three

strokes prevented, one major episode of bleeding would occur. If the lower boundary of the CI for the benefit of oral anticoagulants represents the truth, the NNT is 45 and for every two strokes prevented, one would cause a major episode of bleeding; if, on the other hand, the upper boundary represents the truth, the NNT is 26 and approximately four strokes would be prevented for every major bleeding episode. The true risk-benefit ratio probably lies somewhere between these extremes, closer to that asssociated with the point estimate.

And what about the woman with lupus nephritis, whose plight, described in part A of this two-part essay, prompted us to find a trial of adding plasmapheresis to a regimen of prednisone and cyclophosphamide? Unfortunately, although plasmapheresis did produce sharp declines in the substituted end points of anti-dsDNA antibodies and cryoprecipitable immune complexes, the trial did not find any benefit from plasmapheresis in the clinically important measures of renal failure or mortality. When a careful statistical analysis of the emerging data suggested little hope of ever showing clinical benefit, the trial was stopped.

CONCLUSION

Having read the introduction to this series and the two articles on using articles about therapy, we hope that you are developing a sense of how to use the

medical literature to resolve a treatment decision. First, define the problem clearly, and use one of a number of search strategies to obtain the best available evidence. Having found an article relevant to the therapeutic issue, assess the quality of the evidence. To the extent that the quality of the evidence is poor, any subsequent inference (and the clinical decision it generates) will be weakened. If the quality of the evidence is adequate, determine the range within which the true treatment effect likely falls. Then, consider the extent to which the results are generalizable to the patient at hand, and whether the outcomes that have been measured are important. If the generalizability is in doubt, or the importance of the outcomes questionable, support for a treatment recommendation will be weakened. Finally, by taking into account the patient's risk of adverse events, assess the likely results of the intervention. This involves a balance sheet looking at the probability of benefit and the associated costs (including monetary costs, and issues such as inconvenience) and risks. The bottom line of the balance sheet will guide your treatment decision.

While this may sound like a challenging route to deciding on treatment, it is what clinicians implicitly do each time they administer therapy.[13] Making the process explicit and being able to apply guidelines to help assess the strength of evidence will, we think, result in better patient care.

References

1. Ezekowitz MD, Bridgers SL, James KE, et al, for the Veterans Affairs Stroke Prevention in Nonrheumatic Atrial Fibrillation Investigators. Warfarin in the prevention of stroke associated with nonrheumatic atrial fibrillation. N Engl J Med. 1992; 327:1406-1412.
2. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? JAMA. 1993;270:2598-2601.
3. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. In: Gardner MJ, Altman DG, eds. Statistics With Confidence: Confidence Intervals and Statistical Guidelines. London, England: British Medical Journal; 1989:83-100.
4. Detsky AS, Sackett DL. When was a 'negative'

trial big enough? how many patients you needed depends on what you found. Arch Intern Med. 1985; 145:709-715.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. 2nd ed. Clinical Epidemiology: A Basic Science for Clinical Medicine. Boston, Mass: Little Brown & Co Inc; 1991:218.
6. Guyatt GH, Keller J, Singer J, Halcrow S, Newhouse M. A controlled trial of respiratory muscle training in patients with chronic airflow limitation. Thorax. 1992;47:598-602.
7. Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. Ann Intern Med. 1992;116:78-84.
8. Echt DS, Liebson PR, Mitchell LB, 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.

9. The Cardiac Arrhythmia Suppression Trial II Investigators. Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. N Engl J Med. 1992;327:227-233.
10. Muldoon ME, Manuck SB, Matthews KA. Lowering cholesterol concentrations and mortality: a quantitative review of primary prevention trials. BMJ. 1990;301:309-314.
11. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318: 1728-1733.
12. Levine M, Hirsch J, Ladefeld S, Raskob G. Hemorrhagic complications of anticoagulant treatment. Chest. 1992;102 (suppl):352S-363S.
13. Eddy DM. Clinical policies and the quality of clinical practice. N Engl J Med. 1982;307:343-347.