

MOVING FROM EVIDENCE TO ACTION

Gordon Guyatt, Robert Hayward, W. Scott Richardson,
Lee Green, Mark Wilson, Jack Sinclair, Deborah Cook,
Paul Glasziou, Alan Detsky, and Eric Bass

PJ Devereaux also made substantive contributions to this section

IN THIS SECTION

Finding the Evidence

Treatment Recommendations

Developing Recommendations

Systematic Reviews

Decision Analysis

Practice Guidelines

Assessing Recommendations

Did the Recommendations Consider All Relevant Patient Groups, Management Options, and Possible Outcomes?

Is There a Systematic Review of Evidence Linking Options to Outcomes for Each Relevant Question?

Is There an Appropriate Specification of Values or Preferences Associated With Outcomes?

Do the Authors Indicate the Strength of Their Recommendations?

Are Treatment Recommendations Desirable at All?

Clinical Resolution



CLINICAL SCENARIO

Warfarin in Atrial Fibrillation: Is It the Best Choice for This Patient?

You are a primary care practitioner considering the possibility of warfarin therapy in a 76-year-old woman with congestive heart failure and chronic atrial fibrillation who has just entered your practice. Aspirin is the only antithrombotic agent that the patient has received during the 10 years she has had atrial fibrillation. Her other medical problems include stage I hypertension, which she has had since sometime in her fifth decade, and for which she has been taking hydrochlorothiazide and benazepril. Her previous physicians' records suggest that in recent years her systolic blood pressure was 130 to 140 mm Hg and her diastolic pressure was 80 to 90 mm Hg. Current blood pressure is 136/84 mm Hg, with a heart rate of 76 beats per minute, suggesting effective rate control. The patient does not have valvular disease, diabetes, or other comorbidity, and she does not smoke.

The duration of the patient's atrial fibrillation dissuades you from considering cardioversion or antiarrhythmic therapy. The patient lives alone. Although she has never had a significant fall, you are concerned that warfarin would present a risk of intracranial hemorrhage that may prove to be greater than its benefit in terms of stroke prevention. You find she places a high value on avoiding a stroke and a somewhat lower value on avoiding a major bleeding episode. Although she is not fond of medical care, she would accept the inconvenience associated with monitoring anticoagulant therapy.

The question of whether and when to offer anticoagulant therapy to patients with nonvalvular atrial fibrillation arises often in your practice, but there is little agreement on the topic among you and your partners. You are all convinced that warfarin anticoagulant therapy for nonvalvular atrial fibrillation prevents strokes, but some believe that it causes too many bleeding complications. Several patients in the practice with atrial fibrillation have suffered embolic strokes despite aspirin therapy, but two patients suffered serious gastrointestinal bleeding while taking warfarin. Things became even more confusing recently when one of your colleagues, known as a maverick, declared that clopidogrel is the correct agent to use for patients with nonvalvular atrial fibrillation.

You make no change to the patient's medication regimen today, but you make a note to yourself to reconsider when she returns and to raise the issue at a staff meeting next week.

FINDING THE EVIDENCE

You have little inclination to review the voluminous original literature relating to the benefits of anticoagulant therapy in reducing stroke or its risk of bleeding, but you hope to find an evidence-based recommendation to guide you and your colleagues. You decide to search for two sources of such a recommendation: a practice guideline and a decision analysis.

You bring up your Web browser and go to your favorite search engine, Google.com. Entering the term “practice guidelines,” you see that the second item on the results list is “National Guidelines Clearinghouse,” at www.guidelines.gov. This looks promising, as you note that the server appears to reside at the US Agency for Healthcare Research and Quality (AHRQ), formerly known as the Agency for Health Care Policy and Research (AHCPR), which you recall created a series of guidelines using formal evidence-based guidelines methodology.¹

After linking to the clearinghouse, you see a heading labeled “Guidelines Syntheses.” The syntheses area is described as containing

“ . . . syntheses of selected guidelines that cover similar topic areas.

Key elements of each synthesis include the scope of the guidelines, the interventions and practices considered, the major recommendations and the corresponding rating schemes and strength of the evidence, the areas of agreement, and the areas of disagreement.”

This description seems a close fit for the criteria you have for evidence-based guidelines, but unfortunately, atrial fibrillation is not listed among the syntheses completed thus far. Returning to the main page, you enter the term “atrial fibrillation” in the search box, which yields 22 guidelines. The first one on the list seems promising: “Fifth ACCP Consensus Conference on Antithrombotic Therapy,” from the American College of Chest Physicians, completed in 1998. The guideline is summarized on the Clearinghouse site and has been published in the peer-reviewed literature.² You click on “Complete summary” and then print the text that appears. You also send an e-mail message to the hospital librarian asking for a copy of the published article. You look forward with some trepidation to reading the material, as you are aware that many guidelines, even from sources presumably as authoritative as specialty societies, are poorly constructed.^{3,4}

Before you leave Google.com you enter the phrase “atrial fibrillation decision analysis” in the search text box and the results include the following link:

www.thelancet.com/newlancet/sub/issues/vol355no9208/body.article956.html.

The article is a recent decision analysis published in *The Lancet* that appears highly suitable.⁵



TREATMENT RECOMMENDATIONS

Each day, clinicians make dozens of patient management decisions. Some are relatively inconsequential, whereas others are important. Each one involves weighing benefits and risks, gains and losses, and recommending or instituting a course of action judged to be in the patient's best interest. Implicit in each decision is a consideration of the relevant evidence, an intuitive integration of that evidence, and a weighing of the likely benefits and risks in light of the patient's preferences. When making choices, clinicians may benefit from structured summaries of the options and outcomes, systematic reviews of the evidence regarding the relationship between options and outcomes, and recommendations regarding the best choices. This section of the book explores the process of developing recommendations, suggests how the process may be conducted systematically, and introduces a taxonomy for differentiating recommendations that are more rigorous (and, thus, are more likely to be trustworthy) from those that are less rigorous (and, thus, are at greater risk of being misleading).

Traditionally, authors of original, or primary, research into therapeutic interventions include recommendations about the use of these interventions in clinical practice in the discussion section of their papers. Authors of systematic reviews and meta-analyses also tend to provide their impressions of the management implications of their studies. Typically, however, authors of individual trials or overviews do not consider all possible management options, but instead focus on a comparison of two or three alternatives. They may also fail to identify subpopulations in which the impact of treatment may vary considerably. Finally, when the authors of systematic reviews provide recommendations, they typically are not grounded in an explicit presentation of societal or patient preferences.

Failure to consider these issues may lead to variability in recommendations given the same data. For example, various recommendations emerged from different meta-analyses of selective decontamination of the gut using antibiotic prophylaxis for pneumonia in critically ill patients despite very similar results. The recommendations varied from suggesting implementation, to equivocation, to rejecting implementation.⁶⁻⁹ Varying recommendations reflect the fact that investigators reporting primary studies and meta-analyses often make their recommendations without benefit of an explicit standardized process or set of rules.

When benefits or risks are dramatic and are essentially homogeneous across an entire population, intuition may provide an adequate guide to making treatment recommendations. However, such situations are unusual. In most instances, because of their susceptibility to both bias and random error, intuitive recommendations risk misleading the clinician and the patient.

These considerations suggest that when clinicians examine treatment recommendations, they should critically evaluate the methodologic quality of the recommendations. Our goal in this section is to provide clinicians with the tools to conduct such a critical evaluation.

Although recommendations that impact on health resource allocation may be directed at health policymakers, our focus in this book is to dispense advice for practicing clinicians. We will begin by describing the process of developing a recom-

mentation, and we will introduce two formal processes that clinical investigators, experts, and authoritative bodies use in developing recommendations: decision analysis and clinical practice guidelines. We will then offer criteria for deciding when the process is done well and when it is done poorly, along with a hierarchy of treatment recommendations that clinicians may find useful.

DEVELOPING RECOMMENDATIONS

Figure 1F-1 presents the steps involved in developing a recommendation, along with the formal strategies for doing so. The first step in clinical decision making is to define the decision. This involves specifying the alternative courses of action and the possible outcomes. Often, treatments are designed to delay or prevent an adverse outcome such as stroke, death, or myocardial infarction. As usual, we will refer to the outcomes that treatment is designed to prevent as *target outcomes*. Treatments are associated with their own adverse outcomes—side effects, toxicity, and inconvenience. In addition, new treatments may markedly increase—or decrease—costs. Ideally, the definition of the decision will be comprehensive—all reasonable alternatives will be considered and all possible beneficial and adverse outcomes will be identified. In patients like the woman described in the opening scenario with nonvalvular atrial fibrillation, options include not treating the condition, giving aspirin, or administering anticoagulant therapy with warfarin. Outcomes include minor and major embolic stroke, intracranial hemorrhage, gastrointestinal hemorrhage, minor bleeding, the inconvenience associated with taking and monitoring medication, and costs to the patient, the health care system, and society.

FIGURE 1F-1

A Schematic View of the Process of Developing a Treatment Recommendation

Task	Method for Achieving Task
Specify options and outcomes	Explicit decision framing
↓	
Use evidence to determine the link between options and outcomes in all relevant patient subgroups	Randomized controlled trials and other evidence → Systematic review
↓	
Incorporate values to decide on optimal course of action	Values → Decision analysis or practice guideline
↓	
If necessary, consider local circumstances and modify course of action	Local circumstances → Local guidelines
	Assess local burdens, local barriers, and local resources



Having identified the options and outcomes, decision makers must evaluate the links between the two. What will the alternative management strategies yield in terms of benefit and harm?^{10,11} How are potential benefits and risks likely to vary in different groups of patients?^{11,12} Once these questions are answered, making treatment recommendations involves value judgments about the relative desirability or undesirability of possible outcomes. We will use the term *preferences* synonymously with *values* or *value judgments* in referring to the process of trading off positive and negative consequences of alternative management strategies.

Recently, investigators have applied scientific principles to the identification, selection, and summarization of evidence—and to the valuing of outcomes. We will briefly review the systematic approach to the identification, selection, and summarization of evidence that we have presented in Part 1E, “Summarizing the Evidence,” and will then describe the two strategies used to move from evidence to action—that is, decision analysis and practice guidelines.

Systematic Reviews

Unsystematic approaches to identification and collection of evidence risk biased ascertainment. That is, treatment effects may be underestimated or, more commonly, overestimated, and side effects may be exaggerated or ignored. Even if the evidence has been identified and collected in a systematic fashion, if reviewers are then unsystematic in the way they summarize the collected evidence, they run similar risks of bias. One result of these unsystematic approaches may be recommendations advocating harmful treatment; in other cases, there may be a failure to encourage effective therapy. For example, experts advocated routine use of lidocaine for patients with acute myocardial infarction when available data suggested the intervention was ineffective and possibly even harmful, and they failed to recommend thrombolytic agents when data showed patient benefit.¹³

Systematic reviews deal with this problem by explicitly stating inclusion and exclusion criteria for evidence to be considered, conducting a comprehensive search for the evidence, and summarizing the results according to explicit rules that include examining how effects may vary in different patient subgroups (see Part 1E, “Summarizing the Evidence”). When a systematic review pools data across studies to provide a quantitative estimate of overall treatment effect, we call it a *meta-analysis*. Systematic reviews provide strong evidence when the quality of the primary study design is good and sample sizes are large; they provide weaker evidence when study designs are poor and sample sizes are small. Because judgment is involved in many steps in a systematic review (including specifying inclusion and exclusion criteria, applying these criteria to potentially eligible studies, evaluating the methodologic quality of the primary studies, and selecting an approach to data analysis), systematic reviews are not immune from bias. Nevertheless, in their rigorous approach to identifying and summarizing data, systematic reviews reduce the likelihood of bias in estimating the causal links between management options and patient outcomes.

Decision Analysis

Rigorous *decision analysis* provides a formal structure for integrating the evidence about the beneficial and harmful effects of treatment options with the values or preferences associated with those beneficial and harmful effects. Decision analysis applies explicit, quantitative methods to analyzing decisions under conditions of uncertainty; it allows clinicians to compare the expected consequences of pursuing different strategies. The process of decision analysis makes fully explicit all of the elements of the decision, so that they are open for debate and modification.¹⁴⁻¹⁶

We will use the term *clinical decision analyses* to include studies that use formal, mathematical approaches to analyze decisions faced by clinicians in the course of patient care, such as deciding whether to screen for a condition, choosing a testing strategy, or selecting a type of treatment. Although such analyses can be undertaken to inform a decision for an individual patient (“Should I recommend warfarin to this 76-year-old woman with atrial fibrillation?”), they are undertaken more widely to help inform a decision about clinical policy¹⁷ (“Should I routinely recommend warfarin to patients in my practice with atrial fibrillation?”). The study retrieved by the search in our scenario is an example of the latter, whereas an example of a decision analysis for an individual patient is an analysis of whether to recommend cardiac surgery for an elderly woman with aortic stenosis.¹⁸

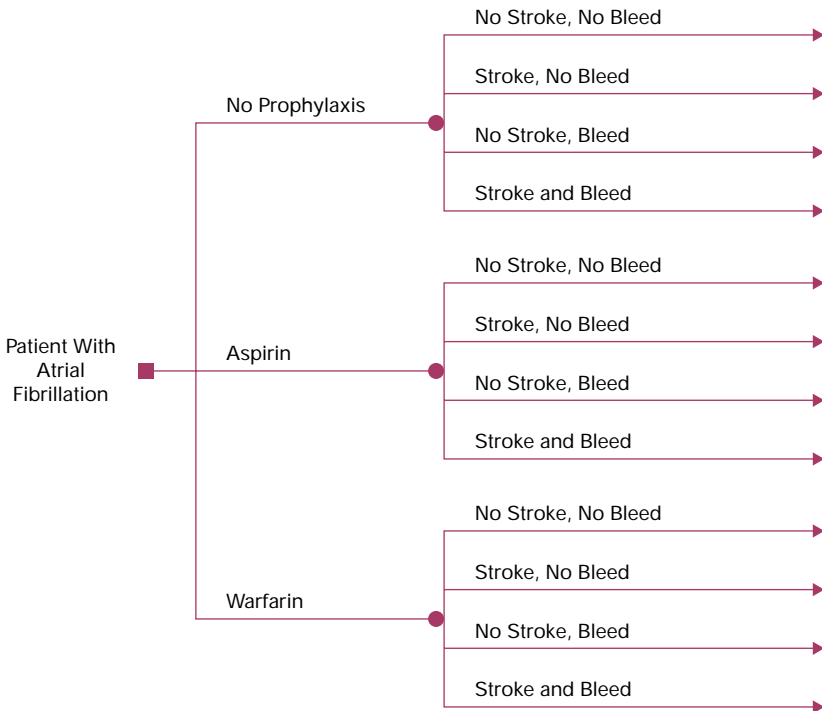
Decision analysis can also be applied to more global questions of health care policy that are viewed from the perspective of society or a national health authority. Examples include analyzing whether or not to screen for prostate cancer¹⁹ and comparing different policies for cholesterol screening and treatment.²⁰ Decision analyses in health services research share many attributes with clinical analyses²¹; however, a discussion of their differences is beyond the scope of this book.

Most clinical decision analyses are built as decision trees, and the articles usually will include one or more diagrams showing the structure of the decision tree used for the analysis. Reviewing such diagrams will help you understand the model. Figure 1F-2 shows a diagram of a very simplified version of the decision tree for the atrial fibrillation problem mentioned at the beginning of this section. The clinician has three options for patients with atrial fibrillation in whom antiarrhythmic therapy to achieve and maintain sinus rhythm is not a possible management strategy: to offer no prophylaxis, to recommend aspirin, or to recommend warfarin. Regardless of what choice is made, patients may or may not develop embolic events and, in particular, stroke. Prophylaxis lowers the chance of embolism but can cause bleeding in some patients. This simplified model excludes a number of important consequences, including the inconvenience of warfarin monitoring and the unpleasantness of minor bleeding.



FIGURE 1F-2

Simplified Decision Tree for a Patient With Atrial Fibrillation



As seen in Figure 1F-2, decision trees are displayed graphically, oriented from left to right, with the decision to be analyzed on the left, the compared strategies in the center, and the clinical outcomes on the right. The decision is represented by a square, termed a *decision node*. The lines emanating from the decision node represent the clinical strategies being compared. Chance events are symbolized by circles, called *chance nodes*, and outcome states are shown (in Figure 1F-2) as triangles or (in other decision trees) as rectangles. When a decision analysis includes costs among the outcomes, it becomes an economic analysis and summarizes tradeoffs between health changes and resource expenditure.^{22, 23} (See Part 2F, "Moving From Evidence to Action, Economic Analysis.")

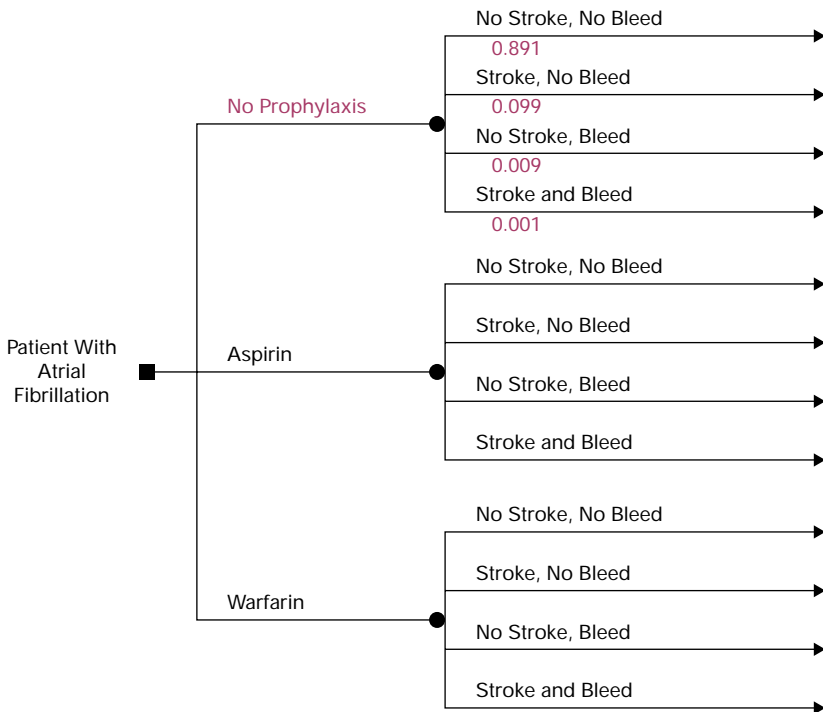
Once a decision analyst has constructed the tree, he or she must generate quantitative estimates of the likelihood of events, or *probabilities*. The scale for probability estimates ranges from 0 (impossible) to 1.0 (absolute certainty). Probabilities must be assigned to each branch emanating from a chance node, and for each chance node, the sum of probabilities must add up to 1.0.

For example, returning to Figure 1F-2, consider the no-prophylaxis strategy (the upper branch emanating from the decision node). This arm has one chance

node at which four possible events could occur (the four possible combinations arising from bleeding or not bleeding and from having a stroke or not having a stroke). Figure 1F-3 depicts the probabilities associated with one arm of the decision, the no-prophylaxis strategy (generated by assuming a 1% chance of bleeding and a 10% probability of stroke, with the two events being independent). Patients given no prophylaxis would have a 0.1% chance (a probability of 0.001) of bleeding and having a stroke, a 0.9% chance (a probability of 0.009) of bleeding and not having a stroke, a 9.9% chance (a probability of 0.099) of not bleeding but having a stroke, and an 89.1% chance (a probability of 0.891) of not bleeding and not having a stroke.

FIGURE 1F-3

Decision Tree With Probabilities—No-Prophylaxis Option



The decision analyst would generate similar probabilities for the other two branches. Presumably, the aspirin branch would have a higher risk of bleeding and a lower risk of stroke. The warfarin branch would have the highest risk of bleeding and the lowest risk of stroke.

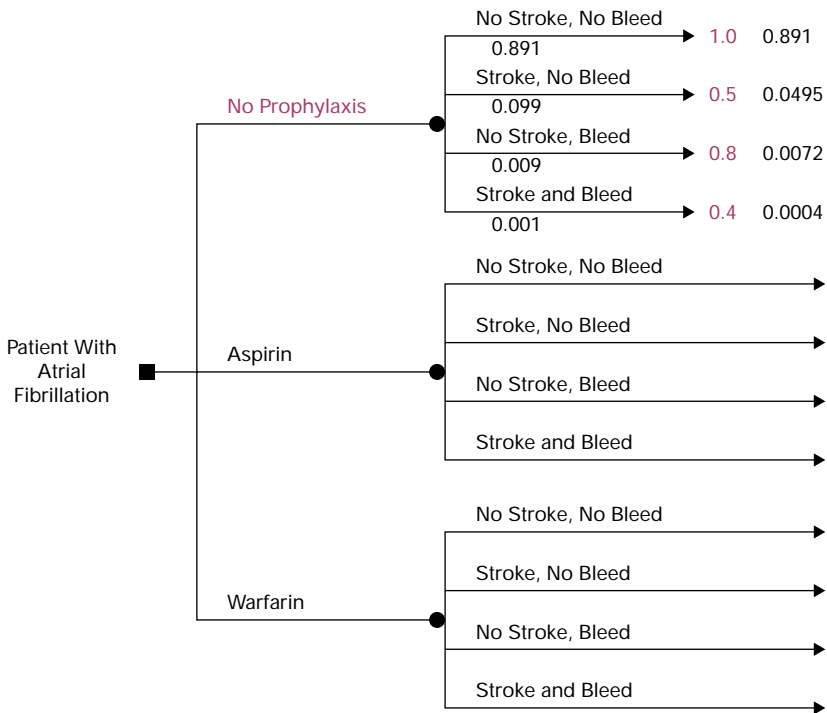
These probabilities would not suggest a clear course of action, as the alternative with the lowest risk of bleeding has the highest risk of stroke, and vice versa. Thus,



the right choice would depend on the relative value or utility one placed on bleeding and stroke. Decision analysts typically place a utility on each of the final possible outcomes that varies from 0 (death) to 1.0 (full health). Figure 1F-4 presents one possible set of utilities associated with the four outcomes and applied to the no-prophylaxis arm of the decision tree: 1.0 for no stroke or bleeding, 0.8 for no stroke and bleeding, 0.5 for stroke but no bleeding, and 0.4 for stroke and bleeding.

FIGURE 1F-4

Decision Tree With Probabilities and Utilities Included in the No-Prophylaxis Arm of the Tree



The final step in the decision analysis is to calculate the total value associated with each possible course of action. Given the particular set of probabilities and utilities we have presented, the value of the no-prophylaxis branch would be $(0.891 \times 1.0) + (0.009 \times 0.8) + (0.099 \times 0.5) + (0.001 \times 0.4)$, or 0.948. Depending on the probabilities attached to the aspirin and warfarin branches, they would be judged superior or inferior to the no-prophylaxis branch. If the total value of each of these branches were >0.948 , they would be judged preferable to the no-prophylaxis branch; if the total value were <0.948 , they would be judged less desirable.

The model presented in Figures 1F-2 to 1F-4 is oversimplified in a number of ways, among which are its omission of the time frame of events and the possibility of a patient suffering multiple events. Decision analysts can make use of software programs that model what might happen to a hypothetical cohort of patients over a series of time cycles (say, periods of 1 year's duration). The model allows for the possibility that patients might move from one health state to another. For instance, one unfortunate patient may suffer a mild stroke in one cycle, continue with minimal functional limitation for a number of cycles, suffer a gastrointestinal bleeding episode in a subsequent cycle and, finally, experience a major stroke. These multistate transition models or Markov models permit more sophisticated and true-to-life depictions—and, presumably, more accurate decision analysis.

Practice Guidelines

Practice guidelines, or “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances,”²⁴ provide an alternative structure for integrating evidence and applying values to reach treatment recommendations.^{1,25-30} Practice guideline methodology places less emphasis on precise quantification than does decision analysis. Instead, it relies on the consensus of a group of decision makers, ideally including experts, front-line clinicians, and patients, who carefully consider the evidence and decide on its implications. The guidelines developers' mandate may be to adduce recommendations for a country, a region, a city, a hospital, or a clinic. Depending on whether the country is the Philippines or the United States, whether the region is urban or rural, whether the institution is a large teaching hospital or a small community hospital, and whether the clinic serves a poor community or an affluent one, guidelines based on the same evidence may differ. For example, clinicians practicing in rural parts of less industrialized countries without resources to monitor its intensity may reject the administration of warfarin to patients with atrial fibrillation.

Both decision analyses and practice guidelines can be methodologically strong or weak and thus may yield either valid or invalid recommendations. In Table 1F-1, we offer four guidelines to assess the validity of a treatment recommendation—one for each step depicted in Table 1F-1—and describe these in detail below.

TABLE 1F-1

Users' Guides for the Validity of Treatment Recommendations

- Did the recommendations consider all relevant patient groups, management options, and possible outcomes?
- Is there a systematic review of evidence linking options to outcomes for each relevant question?
- Is there an appropriate specification of values or preferences associated with outcomes?
- Do the authors indicate the strength of their recommendations?



ASSESSING RECOMMENDATIONS

Did the Recommendations Consider All Relevant Patient Groups, Management Options, and Possible Outcomes?

Recommendations pertain to decisions, and decisions involve particular groups of patients, choices for those patients, and the consequences of the choices. Regardless of whether recommendations apply to diagnosis, prevention, therapy, or rehabilitation, they should specify all relevant patient groups, the interventions of interest, and sensible alternative practices. For example, in a decision analysis of the management of suspected herpes encephalitis, the authors included the three strategies available to clinicians at the time: brain biopsy, empiric vidarabine, or neither option.³¹ Although this model represented the decision well at the time of publication, acyclovir has subsequently become available and is now widely used for this disorder. Because the original model did not include an acyclovir strategy, it would no longer accurately portray the decision.

To cite another example, in a guideline based on a careful systematic literature review,³² the American College of Physicians offers recommendations for medical therapeutic options for preventing strokes.³³ Although the authors mention carotid endarterectomy as an alternative in their practice guidelines, the procedure is not included in the recommendations themselves. These guidelines would have been strengthened if medical management for transient ischemic attacks had been placed in the context of the highly effective surgical procedure.³⁴

Treatment recommendations often vary for different subgroups of patients. In particular, those at lower risk of target outcomes that treatment is designed to prevent are less likely to benefit from therapy than those who are at higher risk (see Part 2B3, “Therapy and Applying the Results, Applying Results to Individual Patients”). For instance, in a guideline concerning hormone replacement therapy in postmenopausal women, the American College of Physicians provided separate recommendations for women who had undergone a hysterectomy and for those at higher risk of cardiovascular disease or breast cancer than for other women.³⁵

Recommendations must consider not only all relevant patient groups and management options, but all important consequences of the options as well. Evidence concerning the effects on morbidity, mortality, and quality of life are all relevant to patients, and efficient use of resources dictates attention to costs. If costs are considered, regardless of whether authors use the perspective of patients, insurers, or the health care system or consider broader issues such as the consequences of time lost from work, they can further affect the conclusions (see Part 2F, “Moving From Evidence to Action, Economic Analysis”). Indeed, a decision analysis that includes economic outcomes is labeled an *economic analysis*.

Making recommendations about screening requires particular attention to identifying all potential outcomes. Attempting to identify disease in asymptomatic individuals may result in a number of negative outcomes that clinicians do not face when diagnosing and treating symptomatic patients. Individuals who screen positive for a disease must live for a longer time with the awareness of their illness and the associated negative psychological consequences. This is particularly

problematic if the condition screened for may remain asymptomatic for long periods of time. For instance, consider a man who screens positive for prostate cancer, but was destined to die of heart disease before the prostate cancer became clinically manifest. Those who screen positive but ultimately prove disease-negative may find the experience traumatic, and people who screen negative but ultimately prove to suffer from the target condition may feel betrayed (see Part 2F, “Moving From Evidence to Action, Recommendations About Screening”).

In their guideline on hormone replacement therapy, the American College of Physicians used lifetime probability of developing endometrial cancer, breast cancer, hip fracture, coronary heart disease, and stroke, along with median life expectancy, to estimate risks and benefits for subgroups of women. They acknowledged possible effects of hormone replacement therapy on serum lipoproteins, uterine bleeding, sexual and urinary function, and the need for invasive monitoring, but they did not include these considerations in the model used to synthesize evidence. The effects of hormone replacement therapy on quality of life, which could have a major impact on patient choices, were not explicitly considered.

In a decision analysis concerning anticoagulant therapy for patients suffering from dilated cardiomyopathy,⁴ the authors’ decision model included all of the clinical events of interest to patients (stroke, other emboli, hemorrhage, etc). The analysts measured outcomes using quality-adjusted life expectancy, a scale that combines information about both the quantity and the quality of life. This metric fit the clinical decision well, for one can expect that warfarin might affect both the quantity and quality of life.

Is There a Systematic Review of Evidence Linking Options to Outcomes for Each Relevant Question?

Having specified options and outcomes, the next task for decision makers is to estimate the likelihood that each outcome will occur. In effect, they have a series of specific questions. For hormone replacement therapy, the initial question is, “what is the effect of alternative approaches on the incidence of hip fracture, breast cancer, endometrial cancer, myocardial infarction, and sudden coronary death?” Recommendations must consolidate and combine all of the relevant evidence in an appropriate manner. In carrying out this task, decision makers must avoid bias that will distort the results. This requires access to, or conduct of, a systematic review of the evidence bearing on each question. Part 1E, “Summarizing the Evidence,” provides guidelines for deciding how likely it is that collection and summarization of the evidence are free from bias.

The best recommendations define admissible evidence, report how it was selected and combined, make key data available for review, and report randomized trials that link interventions with outcomes. However, such randomized trials may be unavailable, and the authors of overviews may reasonably abandon their project if there are no high-quality studies to summarize. Those making recommendations do not have this luxury. For important but ethically, technically, or economically difficult questions, strong scientific evidence may never become available.



Because recommendations must deal with the best (often inadequate) evidence available, a variety of studies (published and unpublished) and reports of expert and consumer experience may need to be considered. This means that the strength of the evidence in support of the recommendations can vary widely. Thus, even recommendations that are grounded in rigorous collection and summarization of evidence may yield weak recommendations if the quality of the evidence is poor, an issue to which we will return later in this section (see Table 1F-1).

Is There an Appropriate Specification of Values or Preferences Associated With Outcomes?

Linking treatment options with outcomes is largely a question of fact and a matter of science. Assigning preferences to outcomes, by contrast, is a matter of values. Consider, for example, the relative importance of a possible increased risk of developing breast cancer compared with expectations of decreased risks for fractures in association with hormone replacement therapy. Consequently, it is important that authors report the principal sources of such judgments and the method of seeking consensus.

Clinicians should look for information about who was involved in assigning values to outcomes or who, by influencing recommendations, was implicitly involved in assigning values. Expert panels and consensus groups are often used to determine what a guideline will say. You need to know who the “experts” are, bearing in mind that panels dominated by members of specialty groups may be subject to intellectual, territorial, and even financial biases. Panels that include a balance of experts in research methodology, practicing generalists and specialists, and public representatives are more likely to have considered diverse views in their deliberations than panels restricted to content area experts.

Even with broad representation, the actual process of deliberation can influence recommendations. Therefore, clinicians should look for a report of methods used to synthesize preferences from multiple sources. Informal and unstructured processes may be vulnerable to undue influence by individual panel members, particularly that of the chair of the panel. Explicit strategies for describing and dealing with dissent among judges, or frank reports of the degree of consensus, strengthen the credibility of the recommendations.

Knowing the extent to which patient preferences were considered is particularly important. Many guideline reports, by their silence on the matter of patient preferences, assume that guideline developers adequately represent patients' interests. Although they are reported rarely, it also would be valuable for you to know which principles—such as patient autonomy, nonmaleficence, or distributive justice—were given priority in guiding decisions about the value of alternative interventions. Excellent guidelines will state whether the guideline is intended to optimize values for individual patients, for reimbursement agencies, or for society as a whole. Ideally, guidelines will state the underlying value judgments on which they are based.

For instance, in the guideline on medical therapies to prevent stroke, the American College of Physicians recommended that aspirin be considered the drug

of choice in patients with transient ischemic attacks and suggested that ticlopidine be reserved for patients who do not tolerate aspirin.²⁴ The best estimate of the effect of ticlopidine relative to aspirin in patients with transient ischemic attacks is a 15% reduction in relative risk, a benefit that would translate into the prevention of one stroke for every 70 patients treated in a group of patients with a 10% risk of stroke. The recommendation that aspirin, rather than ticlopidine, be the drug of choice for patients with transient ischemic attack is made, at least in part, on the basis of the increased cost of ticlopidine and the need for checking the white blood cell count in patients receiving ticlopidine. This implicit value judgment could be questioned, and the guideline would be strengthened if the authors had made explicit the values underlying their judgment.

Clinicians using a decision analysis will not face the huge problem of implicit and hidden value judgments that afflict practice guidelines. The reason, as Figure 1F-4 demonstrates, is that decision analysis requires explicit and quantitative specification of values. These values, expressed as utilities, represent measurements of the value to the decision maker of the various outcomes of the decision. Several methods are available to measure these values directly^{5,7,24,25} (see Part 2B2, “Therapy and Understanding the Results, Quality of Life”); the issue of which of these methods is best remains controversial.

Regardless of the measurement method used, the authors should report the source of the ratings. In a decision analysis built for an individual patient, the most (and probably only) credible ratings are those measured directly from that patient. For analyses built to inform clinical policy, credible ratings could come from three sources. First, they may come from direct measurements from a large group of patients with the disorder in question and to whom results of the decision analysis could be applied. Second, ratings may come from other published studies of quality-of-life judgments by such patients, as was done in a recent analysis of strategies for chronic atrial fibrillation.²⁶ Third, they may come from ratings made by an equally large group of people representing the general public. Whoever provides the rating must understand the outcomes they are asked to rate; the more the raters know about the condition, the more credible are their utility ratings.

Do the Authors Indicate the Strength of Their Recommendations?

Multiple considerations should inform the strength or grade of recommendations: the quality of the sources contributing to the systematic review or reviews that bring together the relevant evidence, the magnitude and consistency of the intervention effects in different studies, the magnitude of adverse effects, the burden to the patient and the health care system, the costs, and the relative value placed upon different outcomes. Thus, recommendations may vary from those that rely on evidence from a systematic review of randomized controlled trials that show large treatment effects on patient-important outcomes with minimal side effects, inconvenience, and costs (yielding a very strong recommendation), to those that rely on evidence from observational studies showing a small magnitude of treatment effect with appreciable side effects and costs (yielding a very weak recommendation).



There are two ways that those developing recommendations can indicate their strength. One, most appropriate for practice guidelines, is to formally grade the strength of a recommendation. The other, most appropriate for decision analyses, is to vary the assumptions about the effect of the management options on the outcomes of interest. In this latter approach, a sensitivity analysis, investigators explore the extent to which varying assumptions might impact the ultimate recommendation. We will discuss the two approaches in turn.

Grades of Recommendation

The Canadian Task Force on the Periodic Health Examination proposed the first formal taxonomy of levels of evidence³⁶⁻³⁸ focusing on individual studies. We have modified this framework, taking into account that practice guidelines must rest on systemic reviews that bring together evidence from the best available individual studies (Table 1F-2).

The letter grades in Table 1F-2 (A, B, C+, and C) reflect a hierarchy of methodologic strength that ranges from overviews of randomized trials with consistent results to overviews of observational studies with inconsistent results. Randomized trials yield the strongest evidence (grade A). Since inferences about the health effects of interventions are weakened when there are unexplained major differences in effects in different studies, guidelines based on randomized trials are stronger when the results of individual studies are similar, and guidelines are weaker when major differences between studies, or heterogeneity, are present (grade B). Recommendations from observational studies yield weaker evidence (grade C).

We now identify two situations in which evidence from RCTs directly addressing the question of interest is unavailable, but the evidence is nevertheless strong. First, generalization from one group of patients to another may be very secure. For instance, randomized trials show a large reduction of strokes in patients with atrial fibrillation without mitral valve disease. The underlying biology suggests that clinicians are on strong ground generalizing these results to patients with atrial fibrillation who do have mitral valve disease. Second, observational studies may yield a very high level of consistency and a very large magnitude of effect. Insulin therapy for acute diabetic ketoacidosis provides an example of such a situation. We denote the strength of evidence in both these contexts as C+.

If the evidence linking interventions and outcomes comes from systematic reviews of original studies, clinicians can apply the criteria for a valid systematic review and the schema in Table 1F-2 to decide on the strength of evidence supporting recommendations.

The number categories in Table 1F-2 (1 and 2) reflect the balance between benefits and risks of therapy. If the benefits clearly outweigh the risks (or vice versa) and virtually all patients would make the same choice, the recommendation is designated grade 1. When the balance is less certain and different patients may make different choices, we designate the recommendation as grade 2. A number of factors may make for uncertainty in the balance between benefits and risks, including marked variation in patient values and a wide range of confidence intervals around estimates of benefit and risk (see Part 2F, "Moving From

Evidence to Action, Grading Recommendations: A Qualitative Approach,” and Part 2F, “Moving From Evidence to Action, Grading Recommendations: A Quantitative Approach”).

TABLE 1F-2

An Approach to Grading Treatment Recommendations Based on Systematic Reviews of the Relevant Evidence

Grade of Recommendation	Clarity of Risk/Benefit	Methodologic Strength of Supporting Evidence	Implications
1 A	Clear	RCTs without important limitations	Strong recommendation; can apply to most patients in most circumstances without reservation
1 B	Clear	RCTs with important limitations (inconsistent results, methodologic flaws*)	Strong recommendations, likely to apply to most patients
1 C+	Clear	No RCTs directly addressing the question, but results from closely related RCTs can be unequivocally extrapolated, or evidence from observational studies may be overwhelming	Strong recommendation; can apply to most patients in most circumstances
1 C	Clear	Observational studies	Intermediate-strength recommendation; may change when stronger evidence is available
2 A	Unclear	RCTs without important limitations	Intermediate-strength recommendation; best action may differ depending on circumstances or patient's or societal values
2 B	Unclear	RCTs with important limitations (inconsistent results, methodologic flaws)	Weak recommendation; alternative approaches likely to be better for some patients under some circumstances
2 C	Unclear	Observational studies	Very weak recommendations; other alternatives may be equally reasonable

* These situations include RCTs with both lack of blinding and subjective outcomes where the risk of bias in measurement of outcomes is high, RCTs with large loss to follow-up.

NOTE: Since grade B and C studies are flawed, it is likely that most recommendations in these classes will be level 2.

The following considerations will bear on whether the recommendation is grade 1 or 2: the magnitude and precision of the treatment effect, patients' risk of the target event being prevented, the nature of the benefit and the magnitude of the risk associated with treatment, variability in patient preferences, variability in regional resource availability and health care delivery practices, and cost considerations. Inevitably, weighing these considerations involves subjective judgment.

RCT indicates randomized controlled trial



If recommendations are developed on the basis of observational studies or if the estimate of the magnitude of the treatment effect is imprecise, clinicians can conclude that the recommendation is relatively weak. Investigators can deal with this weakness in recommendations by testing the effect of the guideline on patient outcomes in a real-world clinical situation. For instance, Weingarten and colleagues examined the impact of implementation of a practice guideline, suggesting that low-risk patients admitted to coronary care units should receive early discharge.³⁹ On alternate months during a 1-year period, clinicians either received or did not receive a reminder of the guideline recommendations. During the months in which the intervention was in effect, hospital stay for coronary care unit patients was approximately 1 day shorter and the average cost was reduced by more than \$1000.00. Mortality and health status at 1 month after discharge were similar in the two groups. Such a study, if methodologically strong, addresses the weakness in the underlying evidence and dramatically raises the grade of the recommendations.

The guideline on hormone replacement therapy described previously demonstrates the limitations of recommendations based on weak evidence.³⁵ Although the guideline did not grade its recommendations, they are based largely on observational studies and would be characterized as 2C in the schema presented in Table 1F-2. In particular, the guideline relied to a large extent on a meta-analysis of observational studies of the impact of hormone replacement therapy on coronary heart disease, suggesting a relative risk reduction of 0.35. Subsequently, in the first large randomized trial in women with established coronary disease, no reduction in coronary events was found with hormone replacement therapy.⁴⁰ Clearly, clinicians should be cautious in their implementation of grade C recommendations.

Sensitivity Analysis

Decision analysts use the systematic exploration of the uncertainty in the data, known as *sensitivity analysis*, to see what effect varying estimates for risks, benefits, and values have on expected clinical outcomes and, therefore, on the choice of clinical strategies. Sensitivity analysis asks the question: is the conclusion generated by the decision analysis affected by the uncertainties in the estimates of the likelihood or value of the outcomes? Estimates can be varied one at a time, termed one-way sensitivity analyses, or can be varied two or more at a time, known as multiway sensitivity analyses. For instance, investigators conducting a decision analysis of the administration of antibiotic agents for prevention of *Mycobacterium avium-intracellulare* in patients with HIV infection found that the cost-effectiveness of prophylaxis decreased if they either assumed a longer life span for patients or made a less sanguine estimate of the drugs' effectiveness.⁴¹ If they simultaneously assumed both a longer life span and decreased drug effectiveness (a two-way sensitivity analysis), the cost-effectiveness decreased substantially. Clinicians should look for a table that lists which variables the analysts included in their sensitivity analyses, what range of values they used for each variable, and which variables, if any, altered the choice of strategies.

Generally, all of the probability estimates should be tested using sensitivity analyses. The range over which they should be tested will depend on the source of the data. If the estimates come from large, high-quality randomized trials with narrow confidence limits, the range of estimates tested can be narrow. The less valid the methods or the less precise the estimates, the wider the range that must be included in the sensitivity analyses.

Decision analysts should also test utility values with sensitivity analyses, with the range of values again determined by the source of the data. If large numbers of patients or knowledgeable and representative members of the general public gave very similar ratings to the outcome states, investigators can use a narrow range of utility values in the sensitivity analyses. If the ratings came from a small group of raters, or if the values for individuals varied widely, then investigators should use a wider range of utility values in the sensitivity analyses. To the extent that the bottom line of the decision analysis does not change with varying probability estimates and varying values, clinicians can consider the recommendation a strong one. When the bottom-line decision shifts with different plausible probabilities or values, the recommendation becomes much weaker.

Table 1F-3 presents a schema for classifying the methodologic quality of treatment recommendations, emphasizing the three key components: consideration of all relevant options and outcomes, a systematic summary of the evidence, and an explicit or quantitative consideration, or both, of societal or patient preferences.

TABLE 1F-3

A Hierarchy of Rigor in Making Treatment Recommendations

Level of Rigor	Systematic Summary of Evidence	Considers All Relevant Options and Outcomes?	Explicit Statement of Values	Sample Methodologies
High	Yes	Yes	Yes	Practice guideline or decision analysis*
Intermediate	Yes	Yes or no	No	Systematic review*
Low	No	Yes or no	No	Traditional review; article reporting primary research

* Sample methodologies may not reflect the level of rigor shown. Exceptions may occur in either direction. For example, if the author of a practice guideline or decision analysis neither systematically collects nor summarizes information and if neither societal nor patients' values are explicitly considered, recommendations will be produced that are of low rigor. Conversely, if the author of a systematic review does consider all relevant options and at least qualitatively considers values, recommendations approaching high rigor can be produced.



Are Treatment Recommendations Desirable at All?

The approaches we have described highlight the view that patient management decisions are always a function of both evidence and preferences. Values are likely to differ substantially among settings. For example, monitoring of anticoagulant therapy might take on a much stronger negative value in a rural setting where travel distances are large, or in a more severely resource-constrained environment where, for example, there is a direct inverse relationship between the resources available for purchase of antibiotic drugs and those allocated to monitoring levels of anticoagulation.

Patient-to-patient differences in values are equally important. The magnitude of the negative value of anticoagulant monitoring, or the relative negative value associated with a stroke vs a gastrointestinal bleeding episode, will vary widely among individual patients, even in the same setting. If decisions are so dependent on preferences, what is the point of recommendations?

This line of argument suggests that investigators should systematically search, accumulate, and summarize information for presentation to clinicians. In addition, investigators may highlight the implications of different sets of values for clinical action. The dependence of the decision on the underlying values—and the variability of values—would suggest that such a presentation would be more useful than a recommendation.

We find this argument compelling. However, its implementation depends on standard methods of summarizing and presenting information that clinicians are comfortable interpreting and using. In addition, it assumes that clinicians will have the time and the methods to ascertain patient values that they can then integrate with the information from systematic reviews of the impact of management decisions on patient outcomes. These requirements are unlikely to be fully met in the immediate future. Moreover, treatment recommendations are likely to remain useful for providing insight, marking progress, highlighting areas where we need more information, and stimulating productive controversy. In any case, clinical decisions are likely to improve if clinicians are aware of the underlying determinants of their actions and are able to be more critical about the recommendations offered to them. Our taxonomy may help to achieve both goals.

CLINICAL RESOLUTION

Let us return to the opening clinical scenario. Addressing the validity of the practice guideline on antithrombotic therapy in atrial fibrillation,² you begin by considering whether the guideline developers have addressed all important patient groups, treatment options, and outcomes. You note that they make separate recommendations for patients at varying risk of stroke, but not for patients at different risk of bleeding. The latter omission may occur because studies of prognosis have been inconsistent in the apparent risk factors for bleeding they identified. You have ruled out antiarrhythmic therapy (which another decision

analysis of which you are aware suggests as the management option of choice⁴²) for the patient before you. The guideline addresses the options you are seriously considering, full- and fixed-dose warfarin and aspirin, but does not mention your eccentric colleague's choice of clopidogrel or a related agent, ticlopidine. The guideline addresses the major outcomes of interest, occlusive (embolic) stroke, hemorrhagic stroke, gastrointestinal bleeding, and other major bleeding events, but does not deal specifically with the need for regular blood testing or the frequent minor bruising associated with warfarin therapy.

Moving to the selection and synthesis of the evidence, you find the guideline's eligibility criteria to be appropriate and the supportive literature search, as documented by the clearinghouse, to be comprehensive. The synthesis method is not stated explicitly, but in reading the text it becomes apparent that it is based on calculation and comparison of absolute and relative event rates for both benefits and risks and that it is tied to the guideline's strength of recommendations.

The authors of the guideline make it clear that they believe patient values are crucial to the decision, although they do not explicitly specify the relative value of stroke and bleeding that underlie their recommendations. The guideline comes down clearly on the side of adjusted-dose warfarin therapy for high-risk patients and aspirin for low-risk patients. Since high-risk patients still bleed with warfarin and low-risk patients experience fewer strokes when they take anticoagulant agents, the recommendations express an implicit relative valuing of strokes vs major and minor bleeding episodes and the inconvenience associated with warfarin therapy.

When, as in this case, guideline developers are implicit, clinicians must examine who the people involved in making recommendations are, and the possible influences on their value judgments. The developers are all expert specialists—the authors do not include patients or primary care physicians. Dupont, the makers of warfarin, funded the production of the guidelines, published as a supplement to the journal *Chest*.² This is worth noting, for the funders of any research project may influence its conduct. When, as is often the case in guidelines, investigators are making implicit value judgments, the possible biases that flow from the source of funding are particularly dangerous.

The guideline developers used the predecessor of the grading scheme described earlier in this section, basing all of their recommendations on the results of randomized controlled trials with consistent results, and thus rated them grade A (see Table 1F-2). They classified both of their recommendations that high-risk patients receive warfarin and low-risk patients aspirin as grade 1, meaning they believe that in both cases, the risk-benefit relationship is clear. The patient from the clinical scenario presented earlier in this section falls into the intermediate-risk category. The recommendations suggest that either warfarin or aspirin represents a reasonable option for her. Overall, the guideline meets validity criteria relatively well, and you are inclined to place a high level of trust in the authors' recommendations.

The decision analysis⁵ restricts its comparison to warfarin therapy vs no treatment. Its rationale for omitting aspirin is that its efficacy is not proven (although the aspirin effect in other meta-analyses has achieved statistical significance, it has



always been on the border). The investigators do not mention any other antiplatelet treatment. They include outcomes of the inconvenience associated with monitoring of anticoagulant therapy, major bleeding episodes, mild stroke, severe stroke, and cost. They omit minor bleeding.

The investigators present their search strategies very clearly. They restrict themselves to the results of computer searches of the published literature but, given this limitation, their searches appear comprehensive. With great clarity, they also describe their rationale for selecting evidence, and their criteria appear rigorous. They note the limitations of one key decision: to choose data from the Framingham study, rather than from randomized controlled trials of therapy for patients with atrial fibrillation, from which to derive their risk estimates.

To generate values, the authors interviewed 57 community-dwelling elderly people with a mean age of 73 years. They used standard gamble methodology (see Part 2B2, "Therapy and Understanding the Results, Quality of Life") to generate utility values. Their key values include utilities, on a 0 to 1.0 scale where 0 is death and 1.0 is full health, of 0.986 for warfarin managed by a general practitioner, 0.880 for a major bleeding episode, 0.675 for a mild stroke, and 0 for a severe stroke.

The investigators conducted a sensitivity analysis that indicated their model was sensitive to variation in patients' utility for being on warfarin. If they assumed utility values for being on warfarin in the upper quartile (1.0; that is, no disutility is suggested for taking warfarin), their analysis suggests that virtually all patients should be receiving warfarin treatment. If they assumed the lower quartile utility, (0.92), the analysis suggests that most patients should not be taking warfarin.

This decision analysis rates high with respect to the validity criteria in Table 1F-1. The utilities in the investigators' core analysis using median patient values and best estimates of risk and risk reduction (their *base case* analysis) match those of the patient in the scenario quite well. The investigators provided tables that suggest the best decision for different patients; when we add the characteristics of the patient being considered in the opening scenario, we find that the verdict is: no benefit from treatment. However, this patient does fit into a cell near the boundary between no benefit and clear benefit, and the investigators' sensitivity analysis suggests that if she places the same value on life taking warfarin as life not taking warfarin, she would benefit from using the drug.

Having reviewed what turns out to be a rigorous guideline and a rigorous decision analysis, you believe that you are in a much stronger position both in your own decision making and in providing guidance to your colleagues. Your residual discomfort stems from the realization that the best decision for many patients, including the patient in the scenario, is critically dependent on the patient's values. You resolve to have a more detailed discussions of the options and the consequences when you see her next (see Part 2F, "Moving From Evidence to Action, Incorporating Patient Values").

References

1. Eddy DM. *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach*. Philadelphia: American College of Physicians; 1992.
2. Laupacis A, Albers G, Dalen J, Dunn MI, Jacobson AK, Singer DE. Antithrombotic therapy in atrial fibrillation. *Chest*. 1998;114:579S-589S.
3. Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA*. 1999;281:1900-1905.
4. Grilli R, Magrini N, Penna A, Mura G, Liberati A. Practice guidelines developed by specialty societies: the need for a critical appraisal. *Lancet*. 2000;355:103-106.
5. Thompson R, Parkin D, Eccles M, Sudlow M, Robinson A. Decision analysis and guidelines for anticoagulant therapy to prevent stroke in patients with atrial fibrillation. *Lancet*. 2000;355:956-962.
6. Vandenbroucke-Grauls CMJE, Vandenbroucke JP. Effect of selective decontamination of the digestive tract on respiratory tract infections and mortality in the intensive care unit. *Lancet*. 1991;338:859-862.
7. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. *BMJ*. 1993;307:525-532.
8. Heyland DK, Cook DJ, Jaeschke R, Griffith L, Lee HN, Guyatt GH. Selective decontamination of the digestive tract. *Chest*. 1994;105:1221-1229.
9. Kollef MH. The role of selective digestive tract decontamination on mortality and respiratory tract infections. *Chest*. 1994;105:1101-1108.
10. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ*. 1995;311:1356-1358.
11. Sinclair JC, Cook R, Guyatt GH, Pauker SG, Cook DJ. When should an effective treatment be used? Derivation of the threshold number needed to treat and the minimum event rate for treatment. *J Clin Epidemiol*. In press.
12. Smith GD, Egger M. Who benefits from medical interventions? *BMJ*. 1994;308:72-74.
13. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. 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.
14. Keeney RL. Decision analysis: an overview. *Operations Res*. 1982;30:803-838.
15. Eckman MH, Levine HJ, Pauker SG. Decision analytic and cost-effectiveness issues concerning anticoagulant prophylaxis in heart disease. *Chest*. 1992;102:538S-549S.



16. Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: a progress report. *Ann Intern Med.* 1987;106:275-291.
17. Eddy DM. Clinical decision making: from theory to practice. Designing a practice policy. Standards, guidelines and options. *JAMA.* 1990;263:3077, 3081, 3084.
18. Wong JB, Salem DN, Pauker SG. You're never too old. *N Engl J Med.* 1993;328:971-974.
19. Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer: a decision analytic view. *JAMA.* 1994;272:773-780.
20. Krahn M, Naylor CD, Basinski AS, Detsky AS. Comparison of an aggressive (U.S.) and a less aggressive (Canadian) policy for cholesterol screening and treatment. *Ann Intern Med.* 1991;115:248-255.
21. Goel V. Decision analysis: applications and limitations. *CMAJ.* 1992;147:413-417.
22. Drummond MF, Richardson WS, O'Brien B, Levine M, Heyland DK, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA.* 1997;277:1552-1557.
23. O'Brien BJ, Heyland DK, Richardson WS, Levine M, Drummond MF, for the Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA.* 1997;277:1802-1806.
24. Institute of Medicine. *Clinical Practice Guidelines: Directions for a New Program.* Washington, DC: National Academy Press; 1990.
25. AMA/Specialty Society Practice Parameters Partnership. *Attributes to Guide the Development of Practice Parameters.* Chicago: American Medical Association; 1990.
26. American College of Physicians. *Clinical Efficacy Assessment Project: Procedural Manual.* Philadelphia: American College of Physicians; 1986.
27. Gottlieb LK, Margolis CZ, Schoenbaum SC. Clinical practice guidelines at an HMO: development and implementation in a quality improvement model. *QRB.* 1990;16:80-86.
28. Lohr KN, Field MJ. A provisional instrument for assessing clinical practice guidelines. 1991. Unpublished.
29. Woolf SH. Expert Panel on Preventive Services: analytic methodology. 1991. Unpublished.
30. Park RE, Fink A, Brook RH, et al. Physicians' rating of appropriate indications for six medical and surgical procedures. *Am J Public Health.* 1986;76:766-772.

31. Barza M, Pauker SG. The decision to biopsy, treat, or wait in suspected herpes encephalitis. *Ann Intern Med.* 1980;92:641-649.
32. Matchar DB, McCrory DC, Barnett HJM, Feussner JR. Medical treatment for stroke prevention. *Ann Intern Med.* 1994;121:41-53.
33. American College of Physicians. Guidelines for medical treatment for stroke prevention. *Ann Intern Med.* 1994;121:54-55.
34. North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N Engl J Med.* 1991;325:445-453.
35. Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med.* 1992;117:1016-1037.
36. Canadian Task Force on the Periodic Health Examination. The periodic health examination. *CMAJ.* 1979;121:1193-1254.
37. Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E. Assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. *J Clin Epidemiol.* 1990;43:891-905.
38. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Arch Intern Med.* 1986;146:464-466.
39. Weingarten SR, Reidinger MS, Conner L, et al. Practice guidelines and reminders to reduce duration of hospital stay for patients with chest pain. *Ann Intern Med.* 1994;120:257-263.
40. Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, Vittinghoff E. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA.* 1998;280:605-613.
41. Bayoumi AM, Redelmeier DA. Preventing *Mycobacterium avium* complex in patients who are using protease inhibitors: a cost-effectiveness analysis. *AIDS.* 1998;12:1503-1512.
42. Catherwood E, Fitzpatrick WD, Greenberg ML, et al. Cost-effectiveness of cardioversion and antiarrhythmic therapy in non-valvular atrial fibrillation. *Ann Intern Med.* 1999;130:625-636.