
Gordon H. Guyatt, Deborah J. Cook, Roman Jaeschke, Stephen G. Pauker and Holger J. Schünemann

Chest 2008;133;123-131
DOI 10.1378/chest.08-0654

The online version of this article, along with updated information and services can be found online on the World Wide Web at:
http://chestjournal.org/cgi/content/abstract/133/6_suppl/123S
Grades of Recommendation for Antithrombotic Agents*

American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition)

Gordon H. Guyatt, MD, FCCP; Deborah J. Cook, MD; Roman Jaeschke, MD; Stephen G. Pauker, MD; and Holger J. Schünemann, MD, PhD, FCCP

This chapter describes the system used by the American College of Chest Physicians to grade recommendations for antithrombotic and thrombolytic therapy as part of the Antithrombotic and Thrombolytic Therapy: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Clinicians need to know if a recommendation is strong or weak, and the methodologic quality of the evidence underlying that recommendation. We determine the strength of a recommendation by considering the balance between the desirable effects of an intervention and the undesirable effects (incremental harms, burdens, and for select recommendations, costs). If the desirable effects outweigh the undesirable effects, we recommend that clinicians offer an intervention to typical patients. The uncertainty associated with the balance between the desirable and undesirable effects will determine the strength of recommendations. If we are confident that benefits do or do not outweigh harms, burden, and costs, we make a strong recommendation in our formulation, Grade 1. If we are less certain of the magnitude of the benefits and risks, burden, and costs, and thus their relative impact, we make a weaker Grade 2 recommendation.

For grading methodologic quality, randomized controlled trials (RCTs) begin as high-quality evidence (designated by “A”), but quality can decrease to moderate (“B”), or low (“C”) as a result of poor design and conduct of RCTs, imprecision, inconsistency of results, indirectness, or a high likelihood for reporting bias. Observational studies begin as low quality of evidence (C) but can increase in quality on the basis of very large treatment effects. Strong (Grade 1) recommendations can be applied uniformly to most patients. Weak (Grade 2) suggestions require more judicious application, particularly considering patient values and preferences and, when resource limitations play an important role, issues of cost.

(CHEST 2008; 133:123S–131S)

Key words: clinical trials; metaanalysis; practice guidelines

Abbreviations: AF = atrial fibrillation; ARR = absolute risk reduction; ASA = acetylsalicylic acid; DVT = deep venous thrombosis; HIT = heparin-induced thrombocytopenia; RCT = randomized controlled trial; RRR = relative risk reduction

The Antithrombotic and Thrombolytic Therapy guideline panel of the American College of Chest Physicians has developed evidence-based guidelines to help clinicians make decisions for typical patients about the prevention or treatment of arterial or venous thromboembolic disease. To integrate these recommendations with their own clinical judgement, clinicians need to understand the basis for the clinical recommendations that expert guidelines offer them. A systematic approach to grading the strength of management recommendations can minimize bias and aid interpretation.

The American College of Chest Physicians has revised its system for grading recommendations used in evidence-based guidelines. This system is being applied to all guidelines for purposes of consistency and comparison with other guidelines. The new grading system is modified from a grading scheme...
developed by the Grades of Recommendation, Assessment, Development and Evaluation Working Group,2,3 which assesses the quality of evidence and strength of recommendations. One advantage of the commonality between the American College of Chest Physicians grading system and Grades of Recommendation, Assessment, Development and Evaluation is that other organizations including Up-ToDate (Waltham, MA), the American College of Physicians, the American Thoracic Society, and the World Health Organization also use one of these schemes.4

Our system of evaluating and presenting recommendations entails an initial assessment of the quality of evidence, followed by judgment about the direction and strength of recommendations. Since clinicians will be most interested in the best course of action, we present the strength of the recommendation first as strong (Grade 1) or weak (Grade 2), followed by the quality of the evidence as high (“A”), moderate (“B”), or low (“C”). Furthermore, we use language for our guidelines that expresses their strength. For strong (Grade 1) recommendations, we say: “We recommend . . . (for or against a particular course of action)”. For weak (Grade 2) recommendations, we say: “We suggest . . . (using or not using)” what we believe to be an optimal management approach. We then specify the methodologic quality with designations of A, B, C. Thus, recommendations can fall into the following categories: 1A, 1B, 1C, 2A, 2B, and 2C (Table 1).

STRENGTH OF THE RECOMMENDATION

In determining the strength of recommendations, our grading system focuses on the degree of confidence in the balance between the desirable and the undesirable effects of an intervention (Table 1). Desirable effects or benefits can include beneficial health outcomes, decreased burden of treatment, and decreased resource use (usually measured as costs). Undesirable effects or downsides can include uncommon major adverse events, common minor side effects, greater burden of treatment, and more resource consumption. Burdens are the demands of adhering to a recommendation that patients or caregivers (eg, family) may dislike, such as taking medication or inconvenient laboratory monitoring or physician visits. If desirable effects of an intervention outweigh undesirable effects, we recommend that clinicians offer the intervention to typical patients. How close the balance between desirable and undesirable effects is, and the uncertainty associated with that balance, will determine the strength of recommendations.

When chapter authors are confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects or vice versa, they make a strong recommendation. Such confidence usually requires high-quality evidence that provides precise estimates of both benefits and downsides, and a clear balance in favor, or against, the benefits vs the downsides of an intervention. On occasion, faced with only low-quality evidence, authors may nevertheless make a strong recommendation. For instance, consider the recommendation for routine monitoring of platelet counts in patients receiving heparin whose risk of heparin-induced thrombocytopenia (HIT) is > 1%. Although the evidence of benefit is weak, the early discontinuation of heparin when the platelet count drops may be of appreciable benefit, and the costs and risks of monitoring are negligible. The authors of the chapter addressing HIT in the 2008 guidelines used this rationale to make a strong recommendation. Similarly, when only low-quality evidence supports an experimental intervention with appreciable costs and/or detriments, authors may recommend strongly against use of that intervention. For instance, authors of the HIT guidelines recommended strongly against routine use of HIT antibody testing, with appreciable cost and risk of false-positive results, in patients without clinical evidence to suggest HIT.

Chapter authors offer a weak recommendation when low-quality evidence results in appreciable uncertainty about the magnitude of benefits and/or downsides, or the benefits and downsides are finely balanced. Other reasons for not being confident include imprecise estimates of benefits or harms, uncertainty or variation in how different individuals value the outcomes and thus their preferences regarding management alternatives, small benefits, or situations when benefits may not be worth the costs (including the costs of implementing the recommendation). While the degree of confidence is a continuum and there is no precise threshold between a strong and a weak recommendation, the presence of important

*From the Departments of Clinical Epidemiology and Biostatistics (Drs. Guyatt and Cook), McMaster University Faculty of Health Sciences, Hamilton, ON, Canada; Polish Institute of Evidence Based Medicine (Dr. Jaeschke), Krakow, Poland; Department of Medicine (Dr. Pauker), Tufts-New England Medical Center, Tufts University School of Medicine, Boston, MA; and Department of Epidemiology (Dr. Schiemann), Italian National Cancer Institute Regina Elena, Rome, Italy.

Manuscript accepted December 20, 2007.

Reproduction of this article is prohibited without written permission from the American College of Chest Physicians (www.chestjournal.org/misc/reprints.shtml).

Correspondence to: Gordon Guyatt, MD, FCCP, McMaster University, Health Sciences Centre, Room 2C12, Hamilton, ON, L8N 3Z5, Canada; e-mail: guyatt@mcmaster.ca

DOI: 10.1378/chest.08-0654

Antithrombotic and Thrombotic Therapy 8th ed: ACCP Guidelines
of between 2% and 40%. We can therefore expect a face risks of death in the first 30 days after infarction failure, typical patients with myocardial infarction their age and factors such as the presence of heart infarction by approximately 25%. Depending on high-quality randomized trial suggest that ASA produces the relative risk of death after myocardial infarction in relation to patient values and preferences. For decisions in which it is clear that benefits far outweigh downsides, or downsides far outweigh benefits, almost all patients will make the same choice, and guideline developers can offer a strong recommendation.

For instance, results from an extremely large high-quality randomized trial suggest that ASA reduces the relative risk of death after myocardial infarction by approximately 25%. Depending on their age and factors such as the presence of heart failure, typical patients with myocardial infarction face risks of death in the first 30 days after infarction of between 2% and 40%. We can therefore expect a 0.5% absolute risk reduction (ARR) [from 2 to 1.5%] in the lowest-risk patients and a 10% ARR (from 40 to 30%) in the highest-risk patients. ASA has minimal side effects and is very inexpensive. Because, even in the lowest-risk subgroups, the benefits clearly outweigh the risks, adverse consequences, and costs, administration of ASA is strongly endorsed and widely used. Using letters and numbers to express the quality of the evidence and strength of recommendations (Table 1), both low-risk and high-risk patients would fall within the category of a strong recommendation based on high-quality evidence or Grade 1A (Grade 1 because it the benefits clearly outweigh the downsides, and A because the estimate of benefit comes from high-quality, randomized trials that yielded consistent results).

Thus, a second way for clinicians to interpret strong recommendations is that they provide, for typical patients, a mandate for the clinician to provide a simple explanation of the intervention along with a suggestion that the patient will benefit from its use. Further elaboration will seldom be necessary; however, when clinicians face weak recommendations, they should more carefully consider the benefits, harms and burden in the context of the patient

### Table 1—Grading Recommendation

<table>
<thead>
<tr>
<th>Grade of Recommendation*</th>
<th>Benefit vs Risk and Burdens</th>
<th>Methodologic Quality of Supporting Evidence</th>
<th>Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong recommendation, high-quality evidence, Grade 1A</td>
<td>Desirable effects clearly outweigh undesirable effects, or vice versa</td>
<td>Consistent evidence from RCTs without important limitations or exceptionally strong evidence from observational studies</td>
<td>Recommendation can apply to most patients in most circumstances; further research is very unlikely to change our confidence in the estimate of effect</td>
</tr>
<tr>
<td>Strong recommendation, moderate-quality evidence, Grade 1B</td>
<td>Desirable effects clearly outweigh undesirable effects, or vice versa</td>
<td>Evidence from RCTs with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies</td>
<td>Recommendation can apply to most patients in most circumstances; higher quality research may well have an important impact on our confidence in the estimate of effect and may change the estimate</td>
</tr>
<tr>
<td>Strong recommendation, low or very low-quality evidence, Grade 1C</td>
<td>Desirable effects clearly outweigh undesirable effects, or vice versa</td>
<td>Evidence for at least one critical outcome from observational studies, case series, or from RCTs with serious flaws or indirect evidence</td>
<td>Recommendation can apply to most patients in many circumstances; higher-quality research is likely to have an important impact on our confidence in the estimate of effect and may well change the estimate</td>
</tr>
<tr>
<td>Weak recommendation, high-quality evidence, Grade 2A</td>
<td>Desirable effects closely balanced with undesirable effects</td>
<td>Consistent evidence from RCTs without important limitations or exceptionally strong evidence from observational studies</td>
<td>The best action may differ depending on circumstances or patient or society values; further research is very unlikely to change our confidence in the estimate of effect</td>
</tr>
<tr>
<td>Weak recommendation, moderate-quality evidence, Grade 2B</td>
<td>Desirable effects closely balanced with undesirable effects</td>
<td>Evidence from RCTs with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies</td>
<td>Best action may differ depending on circumstances or patient or society values; higher-quality research may well have an important impact on our confidence in the estimate of effect and may change the estimate</td>
</tr>
<tr>
<td>Weak recommendation, low or very low-quality evidence, Grade 2C</td>
<td>Desirable effects closely balanced with undesirable effects</td>
<td>Evidence for at least one critical outcome from observational studies, case series, or from RCTs with serious flaws or indirect evidence</td>
<td>Other alternatives may be equally reasonable; higher-quality research is likely to have an important impact on our confidence in the estimate of effect and may well change the estimate</td>
</tr>
</tbody>
</table>

*We use the wording *we recommend* for strong (Grade 1) recommendations and *we suggest* for weak (Grade 2) recommendations.
before them, and ensure that the decision is consistent with the patient’s values and preferences. These situations arise when appreciable numbers of patients, because of variability in values and preferences, will make different choices.

Consider a 40-year-old man who has had an idiopathic deep venous thrombosis (DVT) followed by treatment with adjusted-dose warfarin for 1 year to prevent recurrent DVT and pulmonary embolism. Continuing on standard-intensity warfarin beyond 1 year will reduce his absolute risk for recurrent DVT by >7% per year for several years. The burdens of treatment include taking warfarin daily, monitoring the intensity of anticoagulation with blood tests, living with the increased risk of both minor and major bleeding and, for some, experiencing those events. Patients who are minimally concerned about the lifestyle limitations of taking warfarin, or are particularly concerned about recurrent DVT, would consider the benefits of avoiding DVT worth the downsides of taking warfarin. Other patients are likely to consider the benefit not worth the harms and burden.

Individualization of clinical decision making in weak recommendations remains a challenge. Although clinicians should always consider patient values and preferences, weak recommendations dictate more detailed conversations with patients to ensure that the ultimate decision is consistent with the patient values. For patients who are interested, a decision aid that presents patients with both benefits and downsides of therapy is likely to improve knowledge, decrease decision-making conflict, and may promote a decision most consistent with underlying values and preferences. Using decision aids for strong recommendations, while still potentially helpful for fully informing patients, is less important, and may be inefficient.

Other ways of interpreting strong and weak recommendations relate to performance or quality indicators. Strong recommendations are candidate performance indicators. For weak recommendations, performance could be measured by monitoring whether clinicians have discussed recommended actions with patients or their surrogates or carefully documented the evaluation of benefits and downsides in the patient’s chart. Table 2 summarizes several ways of interpreting antithrombotic recommendations.

Factors That Influence the Strength of a Recommendation

Table 3 shows the factors that chapter authors have considered in deciding on the direction and strength of recommendations they make. The issues in Table 3 are relevant to each of benefits, harms, and burden from therapy.

The methodologic quality of the evidence, which is the first row in Table 3, is a major topic that we will describe later. The second row indicates that our chapter authors have, in general, made stronger recommendations for interventions that decrease adverse outcomes with high patient importance (those to which on average patients assign greater values and preferences) than those that decrease outcomes of lesser patient importance.

The choice of adjusted-dose warfarin vs ASA for prevention of stroke in patients with AF illustrates a number of the factors that will influence the strength of a recommendation. A systematic review and metaanalysis found a relative risk reduction (RRR) of 46% in all strokes with warfarin vs ASA. This large effect supports a strong recommendation for warfarin. Furthermore, the relatively narrow 95% confidence interval (RRR, 29 to 57%) suggests that warfarin provides a RRR of at least 29% and further supports a strong recommendation. At the same time, warfarin is associated with burdens that include keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of both minor and major bleeding. Most patients, however, are much more stroke averse than they are bleeding averse. As a result, almost all patients with high risk of stroke would choose warfarin, suggesting the appropriateness of a strong recommendation.

Table 2—Examples of Implications of Strong and Weak Recommendations

<table>
<thead>
<tr>
<th>Strong recommendation for a particular intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>For patients: most individuals in this situation would want the recommended course of action, and only a small proportion would not. Formal decision aids, while still useful for fully informing patients, will seldom be needed to help individuals make decisions consistent with their values and preferences.</td>
</tr>
<tr>
<td>For clinicians: most individuals should receive the intervention.</td>
</tr>
<tr>
<td>For quality monitors: adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator. If clinicians choose not to follow such a recommendation, they should document their rationale.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weak recommendation for a particular intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>For patients: the majority of individuals in this situation would want the suggested course of action, but many would not.</td>
</tr>
<tr>
<td>For clinicians: decision aids may well be useful helping individuals making decisions consistent with their values and preferences. Examine the evidence or a summary of the evidence yourself and be prepared to discuss that evidence, as well as their values and preferences, with patients.</td>
</tr>
</tbody>
</table>

For quality monitors: clinicians’ discussion of the pros and cons of the intervention with the patients, and their documentation of the discussion, could be used as a quality criterion.
A patient’s baseline risk of the adverse outcome (also called control event risk or rate) that an intervention is expected to prevent may prove a key consideration. Consider another 65-year-old patient with AF and no other risk factors for stroke. This individual’s risk for stroke in the next year is approximately 2%. Dose-adjusted warfarin can, relative to ASA, reduce the risk to approximately 1%. Some patients who are very stroke adverse may consider the down sides of taking warfarin well worth it. Others are likely to consider the benefit not worth the risks and inconvenience. When, across the range of their values and preferences, fully informed patients are liable to make different choices, guideline panels should offer weak (Grade 2) recommendations.

While it is ideal for clinicians to elicit preferences and values directly from patients, and for guideline panels to obtain values and preference estimates from population-based studies, such studies are often unavailable. When value or preference judgments are particularly important for the interpretation of recommendations, chapter authors have made statements about the key values underlying their recommendations.

As benefits and risks become more finely balanced, or more uncertain, decisions to administer an effective therapy also become more cost sensitive. We have considered cost in only a small proportion of the recommendations in which we considered resource issues particularly important.10

### How Methodologic Quality Contributes to Grades of Recommendation

In our grading system, the highest-quality evidence comes from one or more well-designed and well-executed randomized controlled trials (RCTs) yielding consistent and directly applicable results.
High-quality evidence can also come from well-done observational studies yielding very large effects (our guideline for a very large effect is an RRR of at least 80%) [Table 4]. RCTs with important limitations and well-done observational studies yielding large effects constitute the moderate-quality category. Observational studies yielding modest effects, and RCTs with very serious limitations will be rated as low-quality evidence. The remainder of this chapter describes the methodologic quality grading system in more detail.

Factors That Decrease the Quality of Evidence

The following limitations may decrease the quality of evidence supporting a recommendation (Table 5).

Limitation of Methodology: Our confidence in recommendations decreases if studies suffer from major limitations that are likely to result in a biased assessment of the treatment effect. These methodologic limitations include failure to adhere to an intention-to-treat analysis, lack of blinding with subjective outcomes highly susceptible to bias, a large loss to follow-up, or RCTs stopped early for benefit.

For instance, a randomized trial suggests that danaparoid sodium is of benefit in treating HIT complicated by thrombosis.11 That trial, however, was unblinded, and the key outcome was the clinicians’ assessment of when the thromboembolism had resolved, which is a subjective judgment.

Unexplained Heterogeneity of Results (Inconsistent Results): When studies yield widely differing estimates of the treatment effect (heterogeneity or variability in results), investigators should look for explanations for that heterogeneity. For instance, drugs may have larger relative effects in sicker populations or when administered in larger doses. When heterogeneity exists but investigators fail to identify a plausible explanation, the quality of evidence decreases. For example, RCTs12 of pentoxifylline in patients with intermittent claudication have shown conflicting results that so far defy explanation. When patient characteristics can explain the heterogeneity, recommendations for patient subgroups will generally differ.

Indirectness of Evidence: The question being addressed in the guideline differs from the available evidence in regards to the population, intervention, comparison, or outcome. Investigators may have undertaken studies in similar, but not identical populations to those under consideration for a recommendation. For example, many of the antithrombotic therapies rigorously tested in randomized trials in adults are also administered to children. The adult trials provide high-quality evidence for adult recommendations, but because of indirectness they represent only moderate- or low-quality evidence for children. Similarly, recommendations in pregnant women rely on randomized trial results in nonpregnant individuals, and recommendations for patients with artificial valves often rely on trial results from other patient groups at high risk for thrombosis. In each of these situations, evidence quality is rated down for indirectness.

Lack of Precision: When studies include few patients and few events and thus have wide confidence intervals, a guideline panel will judge the quality of the evidence lower than it otherwise would because of resulting uncertainty in the results. For instance, a well-designed and rigorously conducted RCT addressed the use of nadroparin, a low-molecular-weight heparin, in patients with cerebral venous sinus thrombosis.13 Of 30 treated patients, 4 patients had a poor outcome, as did 6 of 29 patients in the control group. The analysis suggests a 7% ARR (which, if true, would correspond to a requirement to treat approximately 14 patients to prevent a single poor outcome), but the confidence interval also included not only a 26% absolute difference in favor of treatment, but also a 12% difference in favor of placebo.

Reporting Bias: The quality of evidence may be reduced if investigators fail to report studies (typically those that show no effect) or outcomes (typically those that may be harmful or for which no

<table>
<thead>
<tr>
<th>Table 4—Underlying Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying Methodology</td>
</tr>
<tr>
<td>RCT and observational studies with very large effects</td>
</tr>
<tr>
<td>Downgraded RCTs or upgraded observational studies</td>
</tr>
<tr>
<td>Observational studies and RCTs with major limitations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5—Factors That May Decrease the Quality of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limitations in the design and implementation of available RCTs suggesting high likelihood of bias</td>
</tr>
<tr>
<td>Inconsistency of results (including problems with subgroup analyses)</td>
</tr>
<tr>
<td>Indirectness of evidence (indirect population, intervention, control, outcomes)</td>
</tr>
<tr>
<td>Imprecision of results (wide confidence intervals)</td>
</tr>
<tr>
<td>High probability of reporting bias</td>
</tr>
</tbody>
</table>
Factors That Increase the Quality of Evidence

Observational studies can provide moderate or strong evidence. While well-done observational studies will generally yield low-quality evidence, there may be unusual circumstances in which guideline panels classify such evidence as moderate or even high quality (Table 6).

Treatment Effect: On rare occasions when methodologically strong observational studies yield large or very large and consistent estimates of the magnitude of a treatment effect, we may be more confident about the results. In those situations, while the observational studies are likely to have provided an overestimate of the true effect, the weak study design may not explain all of the apparent benefit. Thus, despite reservations based on the observational study design we are confident that the effect exists. Table 6 shows how the magnitude of the effect in these studies may move the assigned quality of evidence from low to moderate, or even to high quality.

Factors That May Increase the Quality of Evidence

<table>
<thead>
<tr>
<th>Large magnitude of effect (direct evidence, RR &gt; 2 or RR &lt; 0.5)</th>
<th>No plausible confounders; very large with RR &gt; 5 or RR &lt; 0.2 and no threats to validity</th>
<th>All plausible confounding would reduce a demonstrated effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dose-response gradient</td>
<td></td>
</tr>
</tbody>
</table>

Table 6—Factors That May Increase the Quality of Evidence

Plausible Bias: On occasion, all plausible biases from observational studies may be working to underestimate an apparent treatment effect. For example, if only sicker patients receive an experimental intervention or exposure yet they still fare better, it is likely that the actual intervention or exposure effect is larger than the data suggest.

Dose-Response Gradient: The presence of a dose-response gradient may also increase our confidence in the findings of observational studies and thereby enhance the assigned quality of evidence. For example, our confidence in the result of observational studies that show an increased risk of bleeding in patients who have supratherapeutic anticoagulation levels is increased by the observation that there is association between progressively higher levels of the international normalized ratio and the increased risk of bleeding.

Interpreting the Recommendations

Clinicians, third-party payers, institutional review committees, and the courts should not construe these guidelines as absolute. In general, anything other than a Grade 1A recommendation indicates that the chapter authors acknowledge that other interpretations of the evidence and other clinical policies may be reasonable and appropriate. Even Grade 1A recommendations will not apply to all circumstances and all patients. For instance, we have been conservative in our considerations of cost, and have seldom downgraded recommendations from Grade 1 to Grade 2 on the basis of expense. As a result, in jurisdictions in which resource constraints are severe, alternative allocations may serve the health of the public far better than some of the interventions that we designate Grade 1A. This will likely be true for all lower gross domestic product countries and, with the increasing promotion of expensive drugs with marginal benefits, may be increasingly true for wealthier nations.

Similarly, following Grade 1A recommendations will at times not serve the best interests of patients with atypical values or preferences, or whose risks differ markedly from the usual patient. For instance, consider patients who find anticoagulant therapy extremely aversive, either because it interferes with their lifestyle (e.g., prevents participation in contact sports), or because of the need for monitoring. Clinicians may reasonably conclude that following some Grade 1A recommenda-
tions for anticoagulation for either group of patients will be a mistake. The same may be true for patients with particular comorbidities (eg, a recent GI bleed, repeated falls, or an arteriovenous malformation) or other special circumstances (eg, very advanced age) that put them at unusual risk.

We trust that these observations convey our acknowledgment that no recommendations or clinical practice guidelines can take in to account the often compelling unique features of individual clinical circumstances. No clinician, and no body charged with evaluating a clinician’s actions, should attempt to apply our recommendations in rote or blanket fashion.

**Summary**

The strength of any recommendation depends on two factors: the tradeoff between benefits, risks, burden, and cost, and our confidence in estimates of those benefits and risks. The American College of Chest Physicians grading system classifies the tradeoff between benefits and risks in two categories: strong (Grade 1 recommendations), in which the tradeoff is clear enough that most patients, despite differences in values, would make the same choice; and weak (Grade 2), in which the trade off is less clear, and individual patients values will likely lead to different choices. We grade methodologic strength in three categories: high-quality evidence (A), usually from RCTs; moderate-quality evidence (B), typically from randomized trials with important limitations or observational studies with large effects; and low-quality evidence (C), usually from observational studies. The framework summarized in Table 1 therefore generates recommendations from the very strong (Grade 1A: benefit/risk clear, methods strong) to the very weak (Grade 2C: benefit/risk questionable, methods weak). Whatever the grade of recommendation, clinicians must use their judgment, considering both local and individual patient circumstances and patient values, in helping patients make individual decisions. In general, however, they should place progressively greater weight on expert recommendations as they move from weak recommendations based on low-quality evidence (Grade 2C) to strong recommendations based on high-quality evidence (Grade 1A).

**Conflict of Interest Disclosures**

Dr. Guyatt reveals no real or potential conflicts of interest or commitment.

Dr. Schünemann discloses that in addition to receiving grant monies from Physicians Services Inc Foundation, he has also received unrestricted grants from Leo Pharmaceutical and Bayer. He has served on the speaker bureau of AstraZeneca, Bayer, Boehringer Ingelheim, and Sanofi-Aventis, and has assisted the advisory committees of Boehringer Ingelheim and AstraZeneca.

Dr. Cook discloses that she received grant monies from the Canadian Institutes for Health Research and a dalteparin donation for a peer-review funded trial by the Canadian Institutes for Health Research.

Dr. Jaeschke discloses that he has received honoraria and travel support from AstraZeneca, GlaxoSmithKline, MSD, and Boehringer Ingelheim. He is also the deputy editor of a medical journal that is partially financed by industry advertisement and projects.

Dr. Pauker reveals no real or potential conflicts of interest or commitment.

**References**


