Applying the Grades of Recommendation for Antithrombotic and Thrombolytic Therapy

The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy

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

This article about the grades of recommendation for antithrombotic and thrombolytic therapy is part of the Seventh American College of Chest Physicians Conference on Antithrombotic and Thrombolytic Therapy: Evidence-Based Guidelines. Clinicians need to know whether a recommendation is strong or weak, and about the methodological quality of the evidence underlying that recommendation. We determine the strength of a recommendation by considering the trade-off between the benefits of a treatment, on the one hand, and the risks, burdens, and costs on the other. Here, as elsewhere, we assume that a recommended treatment will increase costs (we recognize this is not always the case, but for simplicity we will continue to make this assumption). If the benefits outweigh the risks, burdens, and costs, we recommend that clinicians offer a treatment to typical patients. The uncertainty associated with the trade-off between the benefits and the risks, burdens, and costs will determine the strength of the recommendations. If we are very certain that the benefits do, or do not, outweigh the risks, burdens, and costs, we make a strong recommendation (in our formulation, Grade 1). If we are less certain of the magnitude of the benefits and the risks, burdens, and costs, and thus of their relative impact, we make a weaker Grade 2 recommendation. We grade the methodological quality of a recommendation according to the following criteria. Randomized clinical trials (RCTs) with consistent results provide evidence with a low likelihood of bias, which we classify as Grade A recommendations. RCTs with inconsistent results, or with major methodological weaknesses, warrant Grade B recommendations. Grade C recommendations come from observational studies or from a generalization from one group of patients included in randomized trials to a different, but somewhat similar, group of patients who did not participate in those trials. When we find the generalization from RCTs to be secure, or the data from observational studies overwhelmingly compelling, we choose a Grade C+. When that is not the case, we designate methodological quality as Grade C.

(CHEST 2004; 126:179S–187S)

Key words: clinical trials; meta-analysis; practice guidelines

Abbreviations: ACCP = American College of Chest Physicians; CI = confidence interval; DVT = deep venous thrombosis; LMWH = low-molecular-weight heparin; MI = myocardial infarction; RCT = randomized clinical trial; tPA = tissue plasminogen activator; UFH = unfractionated heparin

Treatment decisions involve a trade-off between benefits on the one hand, and risks, burdens, and costs on the other. The Consensus Conference on Antithrombotic and Thrombolytic Therapy of the American College of Chest Physicians (ACCP) has developed evidence-based guidelines to help clinicians make treatment decisions for typical patients. To integrate these recommendations with their own clinical judgment, clinicians need to understand the basis for the clinical recommendations that expert guidelines offer to them. A systematic approach to grading the strength of management recommendations can minimize bias and aid interpretation.

The ACCP Consensus Conference on Antithrombotic Therapy has recognized, from the outset, the need for an explicit approach to grading the strength of recommendations. Indeed, the group introduced a formal rating scheme as part of the first consensus conference in 1986[1] and has refined the guidelines over the subsequent five meetings. The formulation that we used in the previous conference, the 6th ACCP Consensus Conference on Antithrombotic Therapy, established grades of recommendation focusing on two aspects of recommendations (Table 1).[2] The first aspect is the trade-off between the benefits of a treatment on the one hand, and the risks, burdens, and costs on the other. If the benefits outweigh the risks, burdens, and costs, experts will recommend that clinicians offer a treatment to typical patients. The uncertainty associated with the trade-off between the benefits and the risks, burdens, and costs will determine the strength of recommendations. If experts are very certain that benefits do, or do not, outweigh risks, burdens, and costs, they will make a strong recommendation (in our formulation, **Grade 1**). If they are less certain of the magnitude of the benefits and the risks, burdens, and costs, and thus their relative impact, they must make a weaker **Grade 2** recommendation.

A second key factor in grading recommendations is the methodological quality of the underlying evidence. Randomized clinical trials (RCTs) with consistent results provide unbiased, **Grade A** recommendations. RCTs with inconsistent results, or with major methodological weaknesses, warrant **Grade B** recommendations. **Grade C** recommendations come from observational studies or from generalization from one group of patients included in randomized trials to a different, but somewhat similar, group of patients who did not participate in those trials. When experts find the generalization from RCTs to be

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

Table 1—*Current Approach to Grades of Recommendations

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

*Since studies in categories B and C 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 **Grade 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 (see Table 2). Inevitably, weighing these considerations involves subjective judgment.

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

secure, or the data from observational studies to be overwhelmingly compelling, they choose a **Grade C+**. In other instances, they choose **Grade C**.

As we have noted in our previous description of the grading system, we express the primacy of the risk/benefit judgment in making recommendations by placing it first with a designation of either 1 or 2. Furthermore, we choose language for our guidelines that expresses their strength. For **Grade 1** guides "We recommend. . . (for or against a particular course of action)." For **Grade 2** guides, "We suggest . . . (using or not using)" what we believe to be an optimal management approach. We then specify the methodological quality with designations of A, B, C, and C+. Thus, recommendations can fall into the following categories (Table 1): 1A; 1C+; 1B; 1C; 2A; 2C+; 2B; and 2C.[2] We note that some may find the idea that a rating of 1C+ is stronger than a rating of 1B counterintuitive. We have retained the C rating in C+ to emphasize that the data come from a source other than directly applicable randomized trials.

We have made one substantive change to the grading system that we presented in 2001.[2] We now the downgrade methodological quality of recommendations in favor of treatments that carry more risk, inconvenience, and cost than the alternatives if sample size is small or event rates are low. Specifically, if the results are not statistically significant ($p > 0.05$ [two-tailed]) or if the addition of a small number of adverse events to the treatment arm would render a result nonsignificant, we downgrade recommendations from otherwise strong randomized trials from **Grade A** to **Grade B**. For those who are less familiar with the approach, we will describe the basis of the grading system in detail. In the remainder of this article, we illustrate how we apply the grades by discussing fundamental principles, including how methodologically strong studies can yield stronger or weaker recommendations depending on the trade-off between the benefits and the risks, burdens, and costs, and how we decide on methodological quality. We then provide examples of each of the possible grades and conclude with some general comments on the interpretation of the grading system.

1.0 How Methodological Quality and Risk/Benefit Contribute to Grades of Recommendations

Both aspirin and thrombolytic agents reduce the relative risk of death after myocardial infarction (MI) by approximately 25%. Depending on their age and factors such as the presence of heart failure, typical patients with MI face risks of death in the first 30 days after infarction of between 4% and 40%. Therefore, we can expect a 1% absolute reduction in probability (*ie*, from 4 to 3%) in the lowest risk patients, and a 10% reduction (*ie*, from 40 to 30%) in the highest risk group. Aspirin has minimal side effects and very low cost. Thrombolytic agents seldom

result in catastrophic bleeding, and streptokinase is only moderately costly. Because, even in the subgroups with the lowest event rates, the benefits clearly outweigh the risks, harms, and costs, the administration of both aspirin and a thrombolytic agent is strongly endorsed and widely practiced.

Consider the following two other treatment choices: whether to administer streptokinase or tissue plasminogen activator (tPA) for thrombolysis in MI; and whether to offer therapy with clopidogrel or aspirin to patients with recent ischemic stroke. Again, because evidence regarding both decisions comes from high-quality RCTs, recommendations will, with respect to methodological rigor, be strong.

The magnitude of the relative risk reduction in mortality with tPA over streptokinase is approximately 12% (the baseline probability is 25% lower, because the comparison is with patients who are already receiving thrombolytic therapy, corresponding to absolute risk reductions of 0.36% [*ie*, 12% of 3%] in low-probability patients and 3.6% [*ie*, 12% of 30%] in high-probability patients), and tPA is associated both with a greater probability of hemorrhagic stroke than is streptokinase and with a substantially greater cost. Here, it is less clear that benefits outweigh risks, harms, and costs, and the recommendation cannot be as strong. The result is the development of inconsistent recommendations by different guideline groups and variable practice. In general, therapy with tPA is preferred over streptokinase in the United States, while European physicians administer streptokinase more frequently than tPA.

Our best estimate is that therapy with clopidogrel reduces the relative risk of subsequent stroke in patients with recent ischemic stroke by approximately 9% (95% confidence interval [CI], 0.3 to 16.5%) relative to aspirin therapy.[3] In a patient with a probability of stroke in the next year of 10%, the 9% relative risk reduction corresponds to an absolute risk reduction of approximately 1% in the next year. However, clopidogrel is far more costly than aspirin (in contrast to thrombolytic agents), must be administered over a long period of time, and, because the lower boundary of the 95% CI could represent the truth, the magnitude of its effect may be close to zero. Thus, despite the reduction in stroke with clopidogrel therapy, many clinicians continue to offer aspirin as the initial treatment of patients with a high probability of ischemic cerebrovascular events.

These examples illustrate how our treatment decisions depend not only on the strength of the methods, but on the balance between the benefits and the risks, burdens, and costs, and on our confidence in that balance.[4] Depending on the balance between the benefits and risks, methodologically strong studies suggesting a benefit of one agent over a placebo or another agent may lead to different recommendations. When side effects are minimal or the patient's probability of experiencing the target event that treatment will prevent is very high, investigators may make a strong recommendation to administer the more effective agent. When the benefits and risks are closely balanced, we may see conflicting recommendations and practice. When risk reductions are small and the probability of toxicity or substantially increased cost is high, investigators may even recommend the less effective agent.

In addition to the risks of treatment, some interventions are associated with a burden of inconvenience or nuisance that results in an appreciable decrement in health-related quality of life. Let us imagine a patient with atrial fibrillation for whom one would recommend receiving dose-adjusted warfarin to prevent an ischemic stroke. The inevitable burdens of the treatment are taking a warfarin pill daily, keeping the dietary intake of vitamin K constant, and monitoring the intensity of anticoagulation therapy with blood tests. When the burden associated with an intervention is high, the balance between the benefits and the risks, burdens, and costs can become less conclusive. In the scenario, the burden associated with monitoring the intensity of anticoagulation therapy can become large if the patients for whom we make a recommendation live in a rural area, are unable to self-monitor the intensity of anticoagulation therapy, and have to travel for several hours to have each blood sample drawn. Thus, balancing the burden against the benefit (which may be a small absolute reduction in the probability of stroke if the patient's baseline risk is low) or against alternative interventions with less burden, such as aspirin therapy, may lead to **Grade 2** recommendations for some patient groups.

As the magnitude of the benefit decreases, and the risks and burdens increase, decisions to administer an effective therapy also become more cost sensitive. While the ACCP conference participants considered cost in deciding on the strength of recommendations, the paucity of rigorous cost-effectiveness analyses, and the wide variability of costs across jurisdictions, led us to take a conservative approach to cost issues. That is, cost considerations influenced the recommendations, and the grades of those recommendations, only when the cost gradient between alternatives appeared to be very large and the marginal benefits of the more expensive therapy appeared to be small.

2.0 The Grades of Recommendations

2.1 Validity, consistency, and generalizability of results

Investigators making treatment recommendations must consider the best estimate of the treatment effect. A rigorous systematic review will yield the strongest evidence, and a meta-analysis pooling data across trials is often appropriate for arriving at the best single estimate of the treatment effect.[5]

Investigators will make their strongest recommendations when their systematic review reveals one or more RCTs yielding consistent results (**Grade A** evidence [Table 1]). When several RCTs yield widely differing estimates of the treatment effect (we label this situation *heterogeneity present*), investigators look for explanations for that heterogeneity. For instance, drugs may have larger relative effects in sicker or in less sick populations. When heterogeneity exists but investigators fail to identify a

plausible explanation, the strength of recommendations from even rigorous RCTs is weaker (**Grade B** evidence [Table 1]). For example, as outlined in the previous consensus conference, RCTs of pentoxifylline in patients with intermittent claudication have shown conflicting results that so far defy definitive explanation.[6] In acknowledging the unexplained heterogeneity, we move recommendations from **Grade A** to **Grade B**.

Our confidence in recommendations also decreases if the available studies are flawed by major deficiencies that are likely to result in a biased assessment of the treatment effect. These severe methodological limitations, which include a very large loss of patients to follow-up or an unblinded study with subjective outcomes that are highly susceptible to bias, lead us to classify studies as **Grade B**. How lack of blinding can influence the grading is exemplified by the recommendation to treat heparin-induced thrombocytopenia complicated by thrombosis with danaparoid sodium. Because the evidence for this recommendation came from an unblinded trial in which the outcome was the clinicians' assessment of when the thromboembolism had resolved, which is a subjective judgment, Hirsh et al[7] assigned **Grade 1B** to this recommendation. Table 2 summarizes the reasons for downgrading the rating of methodological quality of the evidence supporting a recommendation from **Grade A** to **Grade B**.

Because of prognostic differences between groups, and a lack of safeguards such as blinding that can lead to biased ascertainment of outcomes,[8] recommendations based on observational studies are weaker than those from RCTs, whether or not heterogeneity is present (**Grade C** [Table 1]). **Grade C** recommendations also include those in which we extrapolate from randomized trials in one group of patients to a different group of patients, or to similar patients under different circumstances. For example, the previous Antithrombotic Consensus Conference recommended that trauma patients in whom the administration of low-molecular-weight heparin (LMWH) is contraindicated or delayed receive mechanical prophylaxis. This recommendation is based on extrapolation from positive RCTs of intermittent pneumatic compression in other patient groups.[9] We also classify recommendations based on clinical experience as **Grade C**.

The current system for grading recommendations includes a provision for situations in which we are extremely confident about generalization from RCTs or, because of a very large treatment effect, are extremely confident about the results of observational studies. For example, oral anticoagulation therapy in patients with mechanical heart valves has not been compared to placebo in an RCT. However, evidence from observational studies has suggested that the probability of experiencing thromboembolic events without anticoagulation therapy is 12.3% annually in patients with bileaflet prosthetic aortic valves, and higher for other valve types,[10] and that the pooled estimate of the relative risk reduction with oral anticoagulation therapy is 80% (95% CI, 63 to 90%). While the observational studies are likely to overestimate the true effect, the weak study design is very unlikely to explain the entire benefit. Thus, the authors of this article offered a **Grade 1C+** recommendation for the use of oral anticoagulation therapy in patients with mechanical heart valves.

Similarly, investigators have not conducted RCTs of oral anticoagulation therapy for the prevention of ischemic stroke in patients with atrial fibrillation and mitral valve disease. However, a meta-analysis[11] of six randomized trials showed a very large (68%) and precise (95% CI, 50 to 79%) reduction in relative risk with warfarin therapy in patients with nonvalvular atrial fibrillation. Furthermore, the probability of embolism in patients with mitral stenosis and atrial fibrillation is high, and the biology of embolism and of warfarin action in mitral stenosis is very similar in patients with atrial fibrillation with and without mitral stenosis. Therefore, conference participants offer a **Grade**

Table 2—*Reasons for Downgrading Ratings of Methodological Quality from Grade A to B, and Strength of Recommendations From Grade 1 to 2*

Downgrade	Reasons
From **Grade A** to **Grade B**	Small sample size; inconsistency of results; all available studies are of poor quality
From **Grade 1** to **Grade 2**	
Context	We have evidence that on balance suggest that treatment has an impact on a patient-important outcome that would lead most people, in the absence of toxicity (*ie*, side effects), inconvenience, or cost, to receive the treatment. Thus, in the absence of toxicity, inconvenience, or cost, we have a **Grade 1** recommendation to offer treatment. In this situation, we would downgrade from **Grade 1** to **Grade 2** if (1) the toxicity, and/or (2) the inconvenience, and/or (3) the costs, or (4) their collective impact were such that an appreciable number of people would, on balance, find the treatment not worth these downsides.
Impact of small effects of treatment on grade of recommendations	If the effect were so small that, even in the absence of toxicity, inconvenience, or cost, some people would not think that the treatment was worth receiving, it would be a **Grade 2** recommendation, even without any of the downsides. If most people would receive treatment in the absence of toxicity, inconvenience, or cost, we start with a **Grade 1** recommendation. In that case, the smaller the benefit, the smaller the magnitude of toxicity, inconvenience, or cost that is required to move from **Grade 1** to **Grade 2**.

1C+ recommendation for the use of warfarin therapy in patients with mitral stenosis and atrial fibrillation.

2.2 Trading off benefits and risks

When randomized trials provide precise estimates suggesting large treatment effects, and the risks and costs of therapy are small, we can confidently recommend treatment for average patients with compatible values and preferences. We have provided the examples of aspirin therapy and thrombolysis for MI. Another example is the prophylaxis of deep venous thrombosis (DVT) after hip fracture surgery in which heparin or low-intensity oral anticoagulation therapy reduces the probability of DVT by approximately half.[8] Here, because sample sizes of the studies are relatively large and the CIs are sufficiently narrow, and because prophylaxis is associated with low costs and complications, the benefits clearly outweigh the risks, burdens, and costs of prophylaxis, and the recommendation is strong (**Grade 1** [Table 1]). For many recommendations, investigators have not studied the optimal dosing and duration of therapy. These issues are often controversial, and, because of the often limited evidence, authors have not graded dosing and duration for these recommendations.

If the balance between benefits and risks is in doubt, we may have methodologically rigorous studies providing **Grade A** evidence, and recommendations may still be weak (**Grade 2**). We may be uncertain of the magnitude of the benefit, burdens, or costs either because studies have been of poor quality or because sample sizes have been small and estimates imprecise. Alternatively, we may be quite confident of the magnitude of the beneficial and harmful effects but, because they are closely balanced, be uncertain of whether the benefits outweigh the harmful effects. The examples of tPA vs streptokinase therapy for patients after MI, and therapy with clopidogrel in comparison to aspirin for patients with recent ischemic stroke represent such situations. Aspirin therapy for the primary prevention of cardiovascular disease, in which we prevent one to two vascular events by treating 1,000 low-risk subjects for 1 year while causing one to two major bleeding complications, provides another example. **Grade B** or **C** evidence is unlikely to provide accurate estimates of the balance between benefits and risks, therefore the recommendation in these two categories will often be **Grade 2**.

The use of heparin therapy after MI in patients receiving thrombolytic and aspirin therapy provides another example. A systematic review of randomized trials[12] has suggested that in 1,000 patients with infarction treated with heparin 5 fewer patients will die, 3 fewer patients will experience reinfarction, and 1 fewer person will have a pulmonary embolus, while 3 more patients will experience major bleeds. Furthermore, these estimates are not precise, and the advantage in decreased infarctions may be lost after 6 months. The small, imprecise, and possibly transient benefit leaves us less confident about any recommendation to use heparin in this situation. Hence, the recommendation is likely to be **Grade 2**.

These examples illustrate the independent impact of validity and consistency, and the balance of positive and negative impacts of treatment on the strength of recommendations. In situations in which there is doubt about the value of the trade-off, any recommendation will be weaker, moving from **Grade 1** to **Grade 2** (Table 2).

We will be able to make **Grade 1** recommendations only when we have a relatively clear picture of both the benefits and the risks, burdens, and costs, and when the balance between the two clearly favors recommending, or not recommending, the intervention for the typical patient with compatible values and preferences. Table 3 summarizes how a number of factors can reduce the strength of a recommendation, moving it from **Grade 1** to **Grade 2**. Uncertainty about a recommendation to treat may be

Table 3—*Factors That May Weaken a Recommendation to Treat, Changing From Grade 1 to Grade 2**

Issue	Example
Evidence for less serious event than one hopes to prevent	Preventing postphlebitic syndrome with thrombolytic therapy in DVT rather than death from PE
Smaller treatment effect	Clopidogrel vs aspirin leads to a smaller stroke reduction in TIAs (RRR, 8.7%) than anticoagulation therapy vs placebo in AF (RRR, 68%)
Imprecise estimate of treatment effect	ASA vs placebo in AF has a wider confidence interval than ASA for stroke prevention in patients with TIA
Lower risk of target event	Some surgical patients are at very low risk of postoperative DVT and PE, while other surgical patients have considerably higher rates of DVT and PE
Higher risk of therapy	ASA and clopidogrel in acute coronary syndromes have a higher risk for bleeding than ASA alone
Higher costs	TPA has much higher cost than streptokinase in acute MI
Varying values	Most young, healthy people will put a high value on prolonging their lives (and thus incur suffering to do so); the elderly and infirm are likely to vary in the value they place on prolonging their lives (and may vary in the suffering they are ready to experience to do so)
Higher burden of therapy	Taking adjusted-dose warfarin is associated with a higher burden than taking aspirin; warfarin requires monitoring the intensity of anticoagulation therapy and a relatively constant dietary vitamin K intake

*PE = pulmonary embolism; TIA = transient ischemic attack; AF = atrial fibrillation; RRR = relative risk reduction; ASA = acetylsalicylic acid.

introduced if the following conditions apply: (1) the target event that we are preventing is less important (we are more likely to be confident of recommendations to prevent death or stroke than asymptomatic DVT); (2) the magnitude of risk reduction in the overall group is small; (3) the probability of the target event is low in a particular subgroup of patients; (4) the estimate of the treatment effect is imprecise, as reflected in a wide CI around the effect; (5) there is substantial potential harm associated with therapy; or (6) or we expect a wide divergence in values even among average or typical patients. Higher costs would also lead to weaker recommendations to treat. For example, Jackson and Clagett[6] recommended cilostazol therapy for patients with disabling claudication, but, due to the high cost and the resulting uncertain trade-offs among benefit, risk, and cost, the authors downgraded the recommendation to **Grade 2**.

Thinking back to our other examples, the more balanced the trade-off between benefits and risks, the greater the influence of individual patient values in decision making. Virtually all patients, if they understand the benefits and risks, will take aspirin after experiencing a MI or will comply with prophylaxis to reduce the risk of thromboembolism after undergoing hip replacement. Thus, one way of thinking about a **Grade 1** recommendation is that variability in patient values is unlikely to influence treatment choice in average or typical patients.

When the trade-off between benefits and risks is less clear, individual patient values may influence treatment decisions even among patients with average or typical preferences. For example, in considering the duration of anticoagulation therapy after an episode of idiopathic DVT, patients may make different choices depending on the relative values they place on avoiding a fatal pulmonary embolus, on avoiding bleeding, and on the inconvenience and worry associated with repeated testing to determine the intensity of anticoagulation therapy.

When experts or guideline development groups make recommendations, they assume a particular set of values as they weigh the possible beneficial and detrimental outcomes. For the current ACCP guidelines, when value or preference judgments are particularly salient, the ACCP guideline participants have provided the key values that they attached to these outcomes and that influenced the direction of a recommendation or its grade. Participants did not elicit direct or indirect representation from patients in arriving at these values. Moreover, recommendations can only reflect average values, and the guideline developers were aware that when the trade-off between benefits and risks is finely balanced, patients with different values or preferences may make different choices. Thus, **Grade 2** recommendations are those in which variation in patient values or individual physician values will often mandate different treatment choices, even among average or typical patients. An alternative, but similar, interpretation is that a **Grade 2** recommendation suggests that clinicians conduct detailed conversations with patients to ensure that their ultimate recommendation is consistent with the patient's values.

3.0 Examples of Specific Recommendations

Grade 1A recommendations

We have already discussed the administration of aspirin in patients with acute coronary syndromes. Here, there are a number of randomized trials that suggest relative risk reductions in a mortality rate of approximately 25%. The results are strong and consistent, dictating a **Grade A** recommendation. Because aspirin has low toxicity, is convenient to administer, and has low cost, we are confident that the benefits far outweigh the downsides of therapy and offer a **Grade 1A** recommendation.

Methodologically strong randomized trials have consistently failed to show a clear benefit of streptokinase therapy over no thrombolytic therapy in patients with acute ischemic stroke. Specifically, a systematic review and a meta-analysis found no effect of death or dependency at the end of follow-up for streptokinase (odds ratio, 0.94; 95% CI, 0.72 to 1.24). The same trials showed an increase in the number of symptomatic (including fatal) intracranial hemorrhages in patients receiving streptokinase (OR, 5.20; 95% CI, 3.25 to 8.32). With no evidence of benefit and clear evidence of harm, one can make a **Grade 1A** recommendation against the administration of streptokinase.

Because values and preferences bear on decisions about whether a particular recommendation is **Grade 1** or **Grade 2**, there will inevitably be instances in which one could argue for one designation or the other. We have already described how tPA therapy, in comparison to streptokinase therapy, lowers death rates, while increasing stroke rates and cost, in patients who have experienced MI. If one places a substantially higher value on reducing deaths than on causing strokes, values the increased convenience of tPA over streptokinase therapy, and considers the increased cost of tPA to be relatively trivial, one would offer a **Grade 1A** recommendation for tPA therapy over streptokinase therapy. Those who note that patients tend to be very stroke-averse, and that patients are particularly reluctant to receive a therapy that results in a higher stroke risk, would argue for a **Grade 2A** rather than a **Grade 1A** designation.[13,14]

Grade 1C+ recommendations

We have pointed out that, on occasion, one can have extremely compelling evidence of a treatment benefit without a directly relevant RCT. This occurs when observational studies have shown an extremely large effect, or when one can generalize with virtual certainty from indirect evidence. By the term *indirect evidence*, we mean evidence from similar but not identical patients, or similar but not identical therapies. We already have noted an example of the first situation, as follows: observational studies have suggested that the probability of experiencing thromboembolic events without anticoagulation therapy is 12.3% annually in patients with bileaflet prosthetic aortic valves (higher for other valve types)[1] and that the pooled estimate of the relative risk reduction with oral anticoagulation is 80% (95% CI, 63 to 90%). The magnitude of the

effect is sufficiently large that the authors of this author designated the recommendation in favor of anticoagulation as **Grade 1C+**.

Mechanical prophylaxis measures have not been tested in randomized trials in the stroke setting. They have, however, been tested in a wide variety of other settings in which issues of immobility are very similar to those of stroke patients. The primary recommendation for the prophylaxis of patients with thrombotic stroke is with unfractionated heparin (UFH) or LMWH. The recommendation for patients in whom therapy with anticoagulants is contraindicated is mechanical prophylaxis. Because stroke patients are so similar to those who participated in randomized trials that demonstrated the benefits of mechanical methods, there is every reason to think the biology will be similar in such patients.

Grade 1B recommendations

A well-designed and rigorously conducted RCT addressed the use of nadroparin, a LMWH, in patients with cerebral venous sinus thrombosis. Of the 30 patients treated, 3 had a poor outcome, as did 6 of 29 patients in the control group. The investigators' analysis suggested a 38% reduction in the relative risk of a poor outcome, but the result was not statistically significant.

An RCT of therapy with UFH in patients with cerebral sinus thrombosis compared therapy with dose-adjusted UFH to that with placebo in 20 patients with both patients and data collectors blinded to the treatment was stopped early because of a large effect. Eight of 10 patients receiving heparin, and 1 of 10 receiving placebo, experienced a complete recovery.

Relevant evidence also includes an observational study of 43 cerebral sinus thrombosis patients with intracranial bleeding, 27 of whom received dose-adjusted heparin. The mortality rate was 15% in the heparin group compared with 69% in the nonheparin group.

Consider first a recommendation regarding low-molecular weight heparin. One RCT provides direct evidence, and the RCT and the observational study of UFH provide indirect evidence. Given the results of the RCT, we would ordinarily classify the evidence as **Grade A**. Because the results are not statistically significant and we are going to recommend therapy with LMWH, which increases the bleeding risk, we downgrade the recommendation to **Grade B**. Because the authors of this article found the totality of the evidence that LMWH does indeed decrease serious neurologic sequelae to be persuasive, the recommendation becomes **Grade 1B**. An observer who found the evidence less compelling would have given a **Grade 2B** recommendation.

Now consider the recommendation regarding UFH. In this case, the RCT of UFH and the cohort study provided direct evidence, and the LMWH trial provided indirect evidence. While there is an RCT with a statistically significant results, had there been four rather than two treated patients with adverse outcomes event, the result would no longer have been significant. When statistical significance disappears with the addition of just a handful of events in the treatment group, the benefit is much more tenuous. Therefore, we once again downgrade from **Grade A** to **Grade B**. The logic about the trade-off between benefits and risks is similar to that of LMWH resulting in a **Grade 1B** recommendation.

Grade 1C recommendations

Because inferences about treatment effectiveness from observational studies or the use of indirect evidence from RCTs in other populations are weak, we would anticipate few strong (*ie*, **Grade 1**) recommendations arising from **Grade C** evidence. Situations in which therapy has few adverse effects, or in which we are recommending a therapy with less risk than the alternative, provide exceptions.

For instance, consider the decision to use intermittent pneumatic compression or elastic stockings in a patient with an acute stroke and associated immobilization who has a contraindication for therapy with low-dose heparin as prophylaxis for DVT. No RCTs of nonpharmacologic DVT prophylaxis have been performed in these patients. However, RCTs of other populations provide indirect evidence (**Grade C**) about the likely benefits in stroke patients. Despite the relatively weak evidence, the minimal risks of nonpharmacologic prophylaxis mandate a **Grade 1** recommendation for its use.

Patients with trauma are at high risk for venous thromboembolism, but randomized trials have shown that LMWH therapy can lower, but not eliminate, the risk. Some authorities have recommended the use of inferior vena cava filters for trauma patients. There are, however, no randomized trials of their use in this setting, and their incremental effectiveness over LMWH therapy is uncertain. Furthermore, inferior vena cava filters are costly, are associated with short-term and long-term complications, lead to thrombosis at the insertion site, and are associated with the late development of symptomatic DVT. These considerations suggest the inappropriateness of this unproven prophylactic strategy and justify the **Grade 1C** recommendation against its use.

Grade 2A recommendations

As we have pointed out, we may have consistent results from high-quality RCTs available, but if the benefits only slightly outweigh the risks (or vice versa), the resulting recommendation may be **Grade 2**. For instance, four large, methodologically strong RCTs in men without clinically manifest vascular disease consistently show a small absolute benefit in reducing MIs in patients who receive prophylactic aspirin therapy and a very small absolute increase in hemorrhagic strokes, as well as a small increase in GI bleeding. Whether men choose aspirin therapy will depend on their risk of MI (the greater the risk, the more compelling the case for aspirin), and on the values and preferences underlying the choice (the greater the stroke aversion and aversion to GI bleeding, the less compelling the case for aspirin). Thus, the recommendation for aspirin therapy for primary prevention, particularly in the lower risk men, becomes a **Grade 2A** recommendation.

Grade 2B recommendations

Situations in which randomized trials that are seriously flawed or yield inconsistent results, and in which the benefits and risks of alternatives are closely balanced, dictate **Grade 2B** recommendations. Randomized trials of therapy with intra-arterial thrombolysis vs that with thrombectomy in patients with peripheral arterial occlusive disease have shown inconsistent results. For instance, one trial showed a statistically significant lower mortality rate in the group that received thrombolysis, but the mortality rate was similar in others trials. Overall, it appears that the two procedures result in similar rates of limb salvage and probably in similar mortality rates. Thrombolysis reduces the need for major surgical procedures, but at the price of increased bleeding. Intracranial bleeding rates of 1 to 2% present a particularly serious concern. The authors of this article feel that, on balance, most informed patients would choose thrombolysis, and we recommend this alternative. It is clear, however, that many patients (for instance, those who are highly stroke-averse) would prefer thrombectomy. Overall, then, the uncertainty about the best choice in the presence of randomized trials with inconsistent results dictates the **Grade 2B** recommendation.

Grade 2C recommendations

Authors make **Grade 2C** recommendations when only observational studies are available, or when they generalize from patients who were randomized in other populations. Randomized trials suggest that, in patients with middle cerebral artery occlusion, intra-arterial thrombolysis reduces the long-term incidence of death and disability, but the trials included only 220 patients, the results were barely significant, and intracranial hemorrhage rates were substantially higher in treated patients. For patients with basilar artery thrombosis, only observational data are available. Given the almost certain increase in hemorrhage, and the uncertainty of the long-term benefit, ACCP authors designated their recommendation for intra-arterial thrombolysis in basilar artery thrombosis as **Grade 2C**.

There have been no randomized trials of anticoagulant therapy in patients with aortic valve disease. Indirect evidence from randomized trials in patients with prosthetic valves, atrial fibrillation, and coronary artery disease has suggested that warfarin therapy could reduce the risk of atheroembolism in patients with aortic valve disease. The incidence of patient-important embolism appears, however, to be low enough to warrant a recommendation against anticoagulation therapy. Patients who are very concerned about the risk of atheroembolism, and are not troubled by the inconvenience of warfarin therapy or the risk of bleeding, may choose anticoagulation therapy. The indirect evidence limits the recommendation to **Grade C**, and the uncertainty about the balance between benefits and risks dictates a weak **Grade 2** recommendation.

Grade 2C+ recommendations

Experts could choose a **Grade 2C+** for their recommendation if the benefits of an intervention were clear either due to generalization from RCTs or to a very large effect, but the risks of a treatment or cost were also substantial. We did not, however, encounter this situation in grading any recommendations in the current guidelines, but it is plausible that new anticoagulant agents may present such situations.

Interpreting the Recommendations

Clinicians, third-party payers, institutional review committees, or the courts should not construe these guidelines in any way as absolute dictates. In general, anything other than a **Grade 1A** recommendation indicates that the article 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 less industrialized 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 of those whose risks differ markedly from those of the usual patient. For instance, consider patients who find anticoagulant therapy extremely aversive, either because it interferes with their lifestyle (*eg*, prevents participation in contact sports) or because of the need for monitoring. Clinicians may reasonably conclude that following some **Grade 1A** recommendations for anticoagulation therapy for either group of patients will be a mistake. The same may be true for patients with particular comorbidities (*eg*, a recent GI bleed or a balance disorder with repeated falls) 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 into account the often compelling and 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 a rote or blanket fashion.

4.0 Summary

The strength of any recommendation depends on the following two factors: the trade-off between the benefits and the risks, burdens, and costs; and the strength of the methodology that leads us to estimates of the treatment effect. The framework that we used for this and the previous conference captures these factors. We have graded the trade-off between benefits and risks in the two categories: 1, in which the trade-off is clear enough that most patients, despite differences in values, would make

the same choice; and 2, in which the trade-off is less clear, and individual patients values will likely lead to different choices. We grade methodological strength in the following three categories: A, randomized trials, ideally summarized in a meta-analysis, that show consistent results; C+, observational studies with very strong treatment effects or secure generalizations from randomized trials with consistent results; B, randomized trials with inconsistent results; and C, observational studies. The framework summarized in Table 1 therefore generates recommendations from the very strong (**Grade 1A**, benefit/risk clear and methods strong) to the very weak (**Grade 2C**, benefit/risk questionable and methods weak). Whatever the grade of recommendation, clinicians must use their judgment, bringing both local and individual patient circumstances, and patient values to bear in making individual decisions. In general, however, clinicians should place progressively greater weight on expert recommendations as they move from **Grade 2C** to **Grade 1A**.

References

1 Sackett D. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Arch Intern Med 1986; 146:464–465

2 Guyatt G, Schünemann H, Cook D, et al. Grades of recommendation for antithrombotic agents. Chest 2001; 119(suppl): 3S–7S

3 CAPRIE Steering Committee. A randomized, blinded trial of clopidogrel versus aspirin in patients at risk of ischemic events (CAPRIE). Lancet 1996; 348:1329–1339

4 Guyatt G, Hayward R, Richardson WS, et al. Moving from evidence to action. In: Guyatt G, Rennie D, eds. The users' guides to the medical literature: a manual for evidence-based clinical practice. New York, NY: AMA Publications, 2002; 175–200

5 Collins RPR, Baigent C, Sleight P. Aspirin, heparin, and fibrinolytic therapy in suspected acute myocardial infarction. N Engl J Med 1997; 336:847–860

6 Jackson MR, Clagett GP. Antithrombotic therapy in peripheral arterial occlusive disease. Chest 2001; 119(suppl):283S–299S

7 Hirsh J, Warkentin TE, Shaughnessy SG, et al. Heparin and low-molecular-weight heparin: mechanisms of action, pharmacokinetics, dosing, monitoring, efficacy, and safety. Chest 2001; 119(suppl):64S–94S

8 Col NF, Pauker SG. The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? Ann Intern Med 2003; 139:923–929

9 Geerts WH, Heit JA, Clagett GP, et al. Prevention of venous thromboembolism. Chest 2001; 119(suppl):132S–175S

10 Stein PD, Alpert JS, Bussey HI, et al. Antithrombotic therapy in patients with mechanical and biological prosthetic heart valves. Chest 2001; 119(suppl):220S–227S

11 Hart RG, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 1999; 131:492–501

12 Collins R, MacMahon S, Flather M, et al. Clinical effects of anticoagulant therapy in suspected acute myocardial infarction: systematic overview of randomised trials. BMJ 1996; 313:652–659

13 Devereaux PJ, Cox JL, Carter J, et al. Patient preferences for treatment with tissue plasminogen activator (tPA) or streptokinase (SK) [abstract]. Can J Cardiol 2000; 16(suppl):158F

14 Heyland DK, Gafni A, Levine MAH. Do potential patients prefer tissue plasminogen activator (TPA) over streptokinase (SK)? An evaluation of the risks and benefits of TPA from the patient's perspective. J Clin Epidemiol 2000; 53:888–894