
1 A Guide to the Evidence

Holger J. Schünemann

COPYRIGHTED MATERIAL

1

Rating the Quality of Evidence and Making Recommendations

A Guide to the Spectrum of Clinical Research

Holger J. Schünemann, Martin Stanulla, Jan L. Brozek, Gordon G. Guyatt

Introduction

Clinicians require clinical expertise to integrate a patient's circumstances and values with the best-available evidence to initiate decision making in health care (1). Using "best evidence" implies that a hierarchy of evidence exists and that clinicians are more confident about decisions based on evidence that offers greater protection against bias and random error.

Protection against bias and greater confidence in decisions result from high-quality research evidence. We can consider quality of evidence a continuum that reflects the confidence in estimates of the magnitude of effect of alternative patient management interventions on the outcomes of interest. However, gradations of this continuum are useful for communication with practicing clinicians, providing useful summaries of what is known because specific clinical questions aid interpretation of clinical research (see chapter 4).

Aiding interpretation becomes increasingly important considering that much of clinicians' practice is guided by recommendations from experts summarized in clinical practice guidelines and textbooks such as this new book *Evidence-based Hematology*. To integrate recommendations with their own clinical judgment, clinicians need to understand the basis for the clinical recommendations that experts offer them. A systematic approach to grading the quality of evidence and the resulting recommendations for clinicians represent an important step in providing evidence-based recommendations.

In this chapter, we will describe the key features of the "quality of evidence" and how we asked the authors of individual chapters to evaluate the available evidence and formulate their recommendations using a pragmatic approach that falls short of the full development of evidence-based guidelines. Most authors used an approach based on the work of the Grading of Recommendations Assess-

ment, Development, and Evaluation Working Group (GRADE) (2–5). Over 20 international organizations, including the World Health Organization, the American College of Physicians, the American College of Chest Physicians, the American Thoracic Society, the European Respiratory Society, UpToDate®, and the Cochrane Collaboration, are now using the GRADE system.

Question formulation and recommendations in this book

We asked chapter authors to ask clinical questions that are particularly relevant to hematology practice using the framework of identifying the patient population(s), the interventions examined (or exposure), alternative interventions (comparison), and the outcomes of interest (see chapter 4). We then asked them to identify relevant studies related to these questions or sets of questions.

For instance, McRae and Eikelboom asked whether thrombolytic therapy compared with anticoagulant therapy has favorable effects on death, recurrent venous thrombosis, incidence on post-thrombotic syndrome, thrombus lysis, and major bleeding in patients with deep vein thrombosis (see chapter 11).

We also asked the authors to base the answers to their questions on evaluations of the scientific literature, in particular focusing on recent, methodologically rigorous systematic reviews of randomized controlled trials (RCTs). If authors did not identify a recent and rigorous systematic review, they were asked to search for RCTs and summarize the findings of these studies to answer their clinical questions. Observational studies were included only if RCTs did not answer the specific question (or did not provide information on a particular outcome). Thus, the search studies we suggested focused on relevant systematic reviews or meta-analyses (a pooled statistical summary of relevant studies) followed by searches for randomized trials and observational studies if systematic reviews did not exist or did not include sufficient information to answer the posed questions. For example, Imrie and Cheung (chapter 42) searched for systematic reviews and randomized trials in the Cochrane Library (2006, Issue 3) and Medline (1966–August 2006,

Evidence-based Hematology. Edited by Mark A. Crowther, Jeff Ginsberg, Holger J. Schünemann, Ralph M. Meyer, and Richard Lottenberg.
© 2008 Blackwell Publishing, ISBN: 978-1-4051-5747-6.

A Guide to the Evidence

Table 1.1 Grading recommendations.

Grade of recommendation*	Balance of desirable versus undesirable effects	Methodologic quality of supporting evidence
Strong recommendation High-quality evidence 1A	Desirable effects clearly outweigh undesirable effects or vice versa	Consistent evidence from randomized controlled trials without important limitations or exceptionally strong evidence from observational studies.
Strong recommendation Moderate-quality evidence 1B	Desirable effects clearly outweigh undesirable effects or vice versa	Evidence from randomized controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies.
Strong recommendation Low or very low quality evidence 1C	Desirable effects clearly outweigh undesirable effects, or vice versa	Evidence for at least one critical outcome from observational studies, case series, or from randomized controlled trials with serious flaws or indirect evidence.
Weak recommendation High-quality evidence 2A	Desirable effects closely balanced with undesirable effects	Consistent evidence from randomized controlled trials without important limitations or exceptionally strong evidence from observational studies.
Weak recommendation Moderate-quality evidence 2B	Desirable effects closely balanced with undesirable effects	Evidence from randomized controlled trials with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or very strong evidence from observational studies.
Weak recommendation Low or very low quality evidence 2C	Desirable effects closely balanced with undesirable effects	Evidence for at least one critical outcome from observational studies, case series, or from randomized controlled trials with serious flaws or indirect evidence.

*GRADE (Grading of Recommendations Assessment, Development, and Evaluation Working Group) system suggests the use of the wording “we recommend” for strong (Grade 1) recommendations and “we suggest” for weak (Grade 2) recommendations. This grading system is based on the work on the GRADE Working Group. The categories of low and very low quality that GRADE includes in its four category system are collapsed here into a single category, resulting in three categories of quality of evidence.

week 2) on treatment for lymphoma. They identified an outdated systematic review and six RCTs to answer the questions whether patients with limited stage follicular lymphoma should receive systemic therapy in combination with local radiotherapy to improve disease-free survival. They based their answer, in the format of a clinical recommendation, on a summary of the evidence from the six RCTs.

Evaluating the quality of evidence and making recommendations

Many authors applied the GRADE system for evaluating the quality of evidence and for presenting their recommendations. This approach begins with an initial assessment of the quality of evidence, followed by judgments about the direction (for or against) and strength of recommendations. Since clinicians are most interested in the best course of action, the GRADE system usually presents 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), low (C), and very low (D). Authors of this book adopted a version of the grading system that combines the low and very low categories, because for many questions in hematology evidence from RCTs is available. Furthermore, we asked authors to phrase recommendations the way that would express their

strength. For strong (Grade 1) recommendations, many authors chose the words: “We recommend . . . (for or against a particular course of action).” For weak (Grade 2) recommendations, they used the words: “We suggest . . . (using or not using)” what they believed to be an optimal management approach. They then indicated the methodological quality of the supporting evidence labeling them as A (high quality), B (moderate quality), or C (low or very low quality). Thus, recommendations can fall into the following six categories: 1A, 1B, 1C, 2A, 2B, and 2C (Table 1.1).

Strength of the recommendation

In determining the strength of recommendations, the GRADE system focuses on the degree of confidence in the balance between desirable effects of an intervention on the one hand and undesirable effects on the other (Table 1.1). Desirable effects or benefits include favorable health outcomes, decreased burden of treatment, and decreased resource use (usually measured as costs). Undesirable effects, or downsides, include rare major adverse events, common minor side effects, greater burden of treatment, and more resource consumption. We define burdens as the demands of adhering to a recommendation that patients or caregivers (e.g., family) may dislike, such as taking medication, need for inconvenient laboratory monitoring, or physician visits. If desirable effects of

Table 1.2 Determinants of strength of recommendation.

Factors that influence the strength of a recommendation	Comment
Balance between desirable and undesirable effects	A strong recommendation is more likely as the difference between the desirable and undesirable consequences becomes larger. A weak recommendation is more likely as the net benefit becomes smaller and the certainty around that net benefit decreases.
Quality of the evidence	A strong recommendation becomes more likely with higher quality of evidence.
Values and preferences	A strong recommendation is more likely as the variability of or uncertainty about patient values and preferences decreases. A weak recommendation is more likely as the variability or uncertainty about patient values and preferences increases.
Costs (resource allocation)	A weak recommendation is more likely as the incremental costs of an intervention (more resources consumed) increase.

an intervention outweigh undesirable effects, we recommend that clinicians offer the intervention to typical patients. How close is the balance between desirable and undesirable effects and the uncertainty associated with that balance will determine the strength of recommendations.

Table 1.2 describes the factors GRADE relies on to determine the strength of recommendation.

When chapter authors were confident that the desirable effects of adherence to a recommendation outweighed the undesirable effects or vice versa, they offered a strong recommendation. Such confidence usually requires evidence of high or moderate quality that provides precise estimates of both benefits and downsides, and their clear balance in favor of, or against, one of the management options. The authors offered a weak recommendation when low-quality evidence resulted in appreciable uncertainty about the magnitude of benefits or downsides, or the benefits and downsides were finely balanced. We will describe the factors influencing the quality of evidence in subsequent sections of this chapter. Other reasons for not being confident in the balance between desirable and undesirable effects include: (1) imprecise estimates of benefits or harms, (2) uncertainty or variation in how different individuals value particular outcomes and thus their preferences regarding management alternatives, (3) small benefits, or (4) situations when benefits may not be worth the costs (including the costs of implementing the recommendation). Although the balance between desirable and undesirable effects, and thus the strength of a recommendation, is a continuum, the GRADE system classifies recommendations for or against an intervention into two categories: strong or weak. Categorizing recommendations as “strong” or “weak” is inevitably arbitrary. The GRADE Working Group believes that the simplicity and behavioral implications of this explicit grading outweigh the disadvantages.

For instance, the choice of adjusted-dose warfarin versus aspirin for prevention of stroke in patients with atrial fibrillation

exemplifies a number of the factors that influence the strength of a recommendation. A systematic review with meta-analysis found a relative risk reduction (RRR) of 46% in all strokes with warfarin versus aspirin (6). This large effect supports a strong recommendation for warfarin. Furthermore, the fairly narrow 95% confidence interval around this estimate (29% to 57%) suggests that warfarin provides an RRR of at least 29% that further supports 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 bleeding. Most patients, however, are much more stroke averse than they are bleeding averse (7). As a result, most patients with high risk of stroke would choose warfarin, suggesting the appropriateness of a strong recommendation.

A patient’s baseline risk of the adverse outcome (also called control risk or control event rate) that an intervention is expected to prevent can be an important issue. Consider another 65-year-old patient with atrial fibrillation 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 aspirin, reduce the risk to approximately 1%. Some stroke-averse patients may consider the downsides of taking warfarin well worth it. Others are likely to consider the benefit not worth the risks and inconvenience. When fully informed patients are likely to make different choices across the range of their values and preferences, guideline panels should offer weak (Grade 2) recommendations.

While the ideal approach for clinicians is to elicit preferences and values from their patients and to recommend obtaining values and preference estimates from population-based studies, such studies are rarely available. When value or preference judgments are crucial for interpreting recommendations, some chapter authors have made statements about the key values underlying their recommendations.

A Guide to the Evidence

Table 1.3 Implications of strong and weak recommendations.

Implications	Strong recommendation	Weak recommendation
For patients	Most individuals in this situation would want the recommended course of action and only a small proportion would not. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences.	The majority of individuals in this situation would want the suggested course of action, but many would not.
For clinicians	Most individuals should receive the intervention. Adherence to this recommendation according to the guideline could be used as a quality criterion or performance indicator.	Recognize that different choices will be appropriate for different patients and that you must help each patient arrive at a management decision consistent with her or his values and preferences. Decision aids can help individuals making decisions consistent with their values and preferences.
For policy makers	The recommendation can be adapted as policy in most situations	Policy making will require substantial debates and involvement of many stakeholders

For instance, McRae and Eikelboom suggested that clinicians not use thrombolytic therapy routinely in patients with deep venous thrombosis (DVT) (Grade 2B) because this recommendation ascribes a high value to the increased risk of bleeding with thrombolytic therapy.

As benefits and risks become more finely balanced or more uncertain, decisions to administer an effective therapy also become more cost sensitive. We have not asked authors to explicitly include cost in the recommendations, but cost will bear on the implementation of many recommendations in clinical practice (8).

Interpreting strong and weak recommendations

Table 1.3 shows suggestions for interpreting strong and weak recommendations. For decisions in which 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, consistent results from high-quality randomized trials suggest that aspirin reduces the relative risk of death after myocardial infarction by approximately 25%. Depending on age and factors such as the presence of heart failure, typical patients with acute myocardial infarction face risks of death in the first 30 days of between 2% and 40% (9). One can therefore expect a 0.5% absolute reduction in risk (from 2% to 1.5%) in the lowest-risk patients and a 10% reduction (from 40% to 30%) in the highest-risk ones. Aspirin has minimal side effects and is very inexpensive. Because, even in the lowest-risk subgroups, the desirable effects clearly outweigh the undesirable effects, the administration of aspirin is strongly endorsed and widely used. Using letters and numbers to express the quality of the evidence and strength of recommendations (Table 1.1), both low- and high-risk patients would fall within the category of a strong recommendation based on high-quality evidence or Grade 1A (“1” because the desirable effects clearly outweigh the undesirable ones, and “A” because the

evidence comes from high-quality, randomized trials that yielded consistent results).

Therefore, for typical patients, strong recommendations provide a mandate for the clinician to explain 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 consider the benefits, harms, and burden to the patient more carefully and ensure that the decision is consistent with the patient’s values and preferences. These situations arise when appreciable numbers of patients would make different choices because of variability in values and preferences.

Consider a 40-year-old man who has suffered an idiopathic DVT followed by treatment with adjusted-dose warfarin for one year to prevent recurrent DVT and pulmonary embolism. Continuing on standard-intensity warfarin beyond this period will reduce his absolute risk for recurrent DVT by more than 7% per year for several years (10). The burdens of treatment include taking a warfarin pill daily, keeping dietary intake of vitamin K constant, monitoring the intensity of anticoagulation with blood tests, and living with the increased risk of bleeding. Patients who are very averse to a 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 potential harms and burden.

Individualization of clinical decision making in the context of weak recommendations remains a challenge. Although clinicians should always consider patients’ preferences and values, when they face weak recommendations, they should consider more detailed conversations with patients than for strong recommendations to ensure that the ultimate decision is consistent with the patient’s values. A decision aid that presents patients with both benefits and downsides of therapy is likely to improve knowledge, decrease decision-making conflict, and support a decision most consistent with patients’ values and preferences (11). Clinicians cannot use decision aids in all patients because of time constraints and the limited availability of decisions aids. For strong recommendations, the use of decision aids is inefficient.

Table 1.4 Categories of quality of evidence.

Underlying methodology*	Quality rating
RCT and observational studies with very large effects	high
Downgraded RCTs or upgraded observational studies	moderate
Observational studies with control groups & RCTs and with major limitations	low

*RCT, randomized controlled trial.

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.

How methodologic quality of the evidence contributes to strength of recommendation

In the GRADE system, evidence of the highest quality comes from one or more well-designed and well-executed RCTs, yielding consistent and directly applicable results. High-quality evidence can also come from well-done observational studies yielding very large effects (defined as a relative risk reduction of at least 80%) (Table 1.4).

RCTs with important methodologic limitations and well-done observational studies yielding large effects constitute the moderate-quality category. Well-done observational studies yielding modest effects, and RCTs with very serious limitations, will be rated as low-quality evidence. Next, we describe the system of grading the methodologic quality of evidence in more detail.

Factors that decrease the quality of evidence

Table 1.5 shows the limitations may decrease the quality of evidence supporting a recommendation.

1. Limitation of methodology: Our confidence in recommendations decreases if studies suffer from major limitations that are

Table 1.5 Factors that may decrease the quality of evidence.

• Limitations in the design and implementation of available RCTs,* suggesting high likelihood of bias • Inconsistency of results (including problems with subgroup analyses) • Indirectness of evidence (indirect population, intervention, control, outcomes) • Imprecision of results (wide confidence intervals) • High probability of publication bias
--

*RCT, randomized controlled trial.

likely to result in a biased assessment of the treatment effect. These methodologic limitations include lack of blinding when subjective outcomes highly susceptible to bias are measured, failure to adhere to an intention-to-treat principle in the analysis of results, a large loss to follow-up, or stopping the study early because of observed benefit.

For instance, a randomized trial suggests that danaparoid sodium is beneficial in treating heparin-induced thrombocytopenia complicated by thrombosis (12). In that trial, however, there was no blinding, and the key outcome trial was the clinicians’ subjective judgment on when the thromboembolism had resolved.

2. Inconsistent results (unexplained heterogeneity of results): If studies yield widely differing estimates of the treatment effect (heterogeneity or variability in results), investigators should look for explanations for that heterogeneity. For example, interventions may have larger relative effects in sicker populations or when given in larger doses. When heterogeneity exists, but investigators fail to identify a plausible explanation, the quality of evidence decreases. For example, RCTs of pentoxifylline in patients with intermittent claudication have shown conflicting results that defy explanation (13).

3. Indirectness of evidence (i.e., the question addressed in the recommendation is quite different from the available evidence regarding 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 strong evidence for adult recommendations, but because of indirectness, they represent only moderate- or low-quality evidence for children.

4. Imprecision: If studies include few patients and few events and thus have wide confidence intervals, making recommendations includes judging evidence lower than it otherwise would be 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 (14). Of 30 treated patients, 3 had a poor outcome, as did 6 of 29 patients in the control group. The investigators’ analysis suggests a 7% risk difference (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.

5. Publication bias: The quality of evidence can be reduced if investigators fail to report outcomes or selective outcome reporting (typically, those that show no effect) or if other reasons lead to withheld results. Unfortunately, it is often required to make guesses about the likelihood of publication bias.

Factors that increase the quality of evidence

Observational studies can provide moderate or strong evidence (14). Whereas well-done observational studies usually yield low-quality evidence, there may be unusual circumstances in

A Guide to the Evidence

Table 1.6 Factors that may increase the quality of evidence.

- Large magnitude of effect (direct evidence, $RR > 2$ or $RR < 0.5$ with no plausible confounders; very large with $RR > 5$ or $RR < 0.2$ and no threats to validity)
- All plausible confounding would reduce a demonstrated effect
- Dose-response gradient

RR, relative risk.

which guideline panels classify such evidence as moderate or even high quality (Table 1.6).

1. 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 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 1.6 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. For example, a meta-analysis of 37 observational studies evaluating the impact of warfarin prophylaxis in cardiac valve replacement found that the relative risk for thromboembolism with warfarin was 0.17 (95% CI 0.13–0.24). This very large effect suggests a rating of high-quality evidence (16).

2. 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.

3. 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 of a dose-response gradient between higher levels of the international normalized ratio (INR) and the increased risk of bleeding (17).

Interpreting the recommendations

Clinicians, third-party payers, institutional review committees, and the courts should not construe recommendations in this book 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 patients in all circumstances, and 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 monitoring is needed. Clinicians may reasonably conclude that following some Grade 1A recommendations for anticoagulation for either group of patients will be a mistake. The same may be true for patients with particular comorbidities (e.g., a recent gastrointestinal bleed, repeated falls, or an arteriovenous malformation) or other special circumstances (e.g., very advanced age) that put them at unusual risk. No clinician, and nobody charged with evaluating clinician’s actions, should attempt to apply the recommendations in rote or blanket fashion.

Summary

The strength of any recommendation for practice depends on two factors: the trade-off between desirable factors and undesirable factors (risks, burden, and cost) and our confidence in estimates of those effects. The GRADE framework, with the minor modifications adopted by the authors of this book, classifies the trade-off between desirable and undesirable effects in 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. Three categories of methodologic strength exist: (A) high-quality evidence, usually from RCTs; (B) randomized trials with important limitations or observational studies with large effects; and (C) usually from observational studies. The framework summarized in Table 1.1 therefore generates recommendations from the very strong (1A: desirable and undesirable effects clear, methods high quality) to the very weak (2C: desirable and undesirable effects questionable, methods low quality). Clinicians must use their judgment when applying the recommendations, considering both local and individual patient circumstances and patient values, to help patients make individual decisions. In general, however, clinicians should place progressively greater weight on expert recommendations as they move from 2C to 1A.

References

- 1 Haynes RB, Devereaux PJ, Guyatt GH. Physicians’ and patients’ choices in evidence based practice. *BMJ*. 2002;**324**(7350):1350.
- 2 Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;**328**(7454):1490.
- 3 Guyatt G, Gutterman D, Baumann M, et al. Grading strength of recommendations and quality of evidence in clinical guidelines. *Chest*. 2006;**129**:174–81.
- 4 Guyatt G, Vist G, Falck-Ytter Y, et al. An emerging consensus on grading recommendations? (editorial) *ACP J Club*. 2006;**144**(1):A8.
- 5 Schunemann HJ, Jaeschke R, Cook DJ, et al. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med*. 2006;**174**(5):605–14.

Chapter 1 Rating the Quality of Evidence

6

van Walraven C, Hart RG, Singer DE, et al. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. *JAMA*. 2002;**288**(19):2441–48.

7

Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ*. 2001;**323**(7323):1218–22.

8

Guyatt G, Baumann M, Pauker S, et al. Addressing resource allocation issues in recommendations from clinical practice guideline panels: suggestions from an American College of Chest Physicians task force. *Chest*. 2006;**129**(1):182–87.

9

Stevenson R, Ranjadayalan K, Wilkinson P, et al. Short and long term prognosis of acute myocardial infarction since introduction of thrombolysis. *BMJ*. 1993;**307**:349–53.

10

Büller HR, Agnelli G, Hull RD, et al. Antithrombotic therapy for venous thromboembolic disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;**126**(3 Suppl):401S–28S.

11

O'Connor AM, Stacey D, Entwistle V, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database Syst Rev*. 2003(2):CD001431.

12

Chong BH, Gallus AS, Cade JF, et al. Prospective randomised open-label comparison of danaparoid with dextran 70 in the treatment of heparin-induced thrombocytopenia with thrombosis: a clinical outcome study. *Thromb Haemost*. 2001;**86**(5):1170–75.

13

Clagett GP, Sobel M, Jackson MR, et al. Antithrombotic therapy in peripheral arterial occlusive disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;**126**(3 Suppl):609S–26S.

14

de Bruijn SF, Stam J. Randomized, placebo-controlled trial of anticoagulant treatment with low-molecular-weight heparin for cerebral sinus thrombosis. *Stroke*. 1999;**30**(3):484–88.

15

Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;**334**(7589):349–51.

16

Cannegieter SC, Rosendaal FR, Briet E. Thromboembolic and bleeding complications in patients with mechanical heart valve prostheses. *Circulation*. 1994;**89**(2):635–41.

17

Levine MN, Raskob G, Beyth RJ, et al. Hemorrhagic complications of anticoagulant treatment: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. *Chest*. 2004;**126**(3 Suppl): 287S–310S.