CLINICAL PRACTICE GUIDELINES

CLINICAL PRACTICE GUIDELINES: FACILITATING THEIR USE AND ENHANCING THEIR TRUSTWORTHINESS

By IGNACIO NEUMANN, M.D. M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2014)
Hamilton, Ontario (Health Research Methodology)

TITLE: Clinical Practice Guidelines: Facilitating Their Use and Enhancing Their Trustworthiness

AUTHOR: Ignacio Neumann, M.D. (Pontificia Universidad Católica de Chile) M.Sc. (McMaster University)

SUPERVISOR: Professor Gordon Guyatt

NUMBER OF PAGES: xi, 114

Abstract

Clinicians in general value the use of the best evidence in decision-making and consider that can improve patient care. However, a successful evidence based practice is hard to achieve in real life. In recent years, with the consolidation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, the development of improved standards to judge the trustworthiness of guideline recommendations and the adoption of more strict policies to limit the influence of conflict of interests, trustworthy guidelines have become an attractive alternative for an evidence-based clinical practice.

In this thesis we offer an explicit and easy-to-use guidance to clinicians regarding how to use guideline recommendations in the context of a real life practice. We also provide an in-depth explanation of the judgments involved in determining the direction and strength of recommendations. Finally, we expand the knowledge about how to manage conflict of interests in guideline developers. Through two studies evaluating the conflict of interest policy implemented at the American College of Chest Physicians 9th edition of the Antithrombotic Guidelines, we show what aspects of the policy were successful and what aspects need to be reformulated.

Preface

This thesis is a "sandwich thesis" consisting of four individual manuscripts. At the time of writing (June, 2014) two of the four individual manuscripts (chapters 4 and 5) have been published in peer reviewed journals, and the remaining two (chapters 2 and 3) have been accepted for publication and will be included in the upcoming third edition of the Journal of American Medical Association "Users' guide to medical literature".

Dr. Neumann's contribution to this work include: developing the research questions, designing the studies, writing the protocol, performing the data extraction, conducting the analyses, writing up all manuscripts, submitting the manuscripts and responding to reviewers' comments.

Acknowledgments:

This life changing experience would have not been possible without my supervisor and thesis committee members Dr. Gordon Guyatt, Dr. Holger Schünemann and Dr. Eli Akl. I would like to express the deepest appreciation to them, for their constant support and for the amazing opportunities they have offered me. No words can express my gratitude.

Table of contents

Lists of Figures and Tables

CHAPTER 2:

CHAPTER 3:

CHAPTER 4:

CHAPTER 5:

CHAPTER 6:

List of Abbreviations

ACCP: American College of Chest Physicians

AHA: American Heart Association

ARIA: Allergic Rhinitis and its Impact on Asthma

ARS: Audience Response System

ASA: Acetylsalicylic Acid

AT9: American College of Chest Physicians 9th edition of the Antithrombotic Guidelines

CHADS2: Clinical prediction rule for estimating the risk of stroke in patients with nonvalvular atrial fibrillation

CI: Confidence Interval

COI: Conflicts of Interest

COPD: Chronic Obstructive Pulmonary Disease

GRADE: Grading of Recommendations Assessment, Development and Evaluation

HIT: Heparin-induced thrombocytopenia

HR: Hazard Ratio

INR: International Normalized Ratio

IPCD: Intermittent Pneumatic Compression Device

LMWH: Low-Molecular-Weight Heparin

MI: Myocardial Infarction

NA: Not available

PE: Pulmonary Embolism

RCT: Randomized Clinical Trials

RR: Risk Ratio

SoF: Summary-of-Findings

USPSTF: United States Preventive Services Task Force

VKA: Vitamin K Antagonist

VTE: Venous Thromboembolism

WHO: World Health Organization

HIT: Heparin-induced thrombocytopenia.

Declaration of Academic Achievement

Dr. Neumann was the main contributor and first author of all the manuscripts presented in this thesis. The details of the contributions of coauthors are described at the end of each manuscript.

Chapter 1: Introduction

Clinicians in general value the use of the best evidence in decision-making and consider that can improve patient care.[1-3] However, a successful evidence based practice is hard to achieve.[4] The overwhelming amount of information scattered among different sources[5] and the typically challenging presentation of research findings[6,7] make it virtually impossible to identify, assess and understand the relevant evidence in a timely fashion.

The use of systematic reviews and evidence summaries is an alternative to bridge the gap between research evidence and decision-makers.[8] However, nowadays, with a production of 11 systematic review per day,[9] widely scattered among different journals,[10] systematic reviews do not solve the problem of an unmanageable amount of information. Further, with recent developments, like multiple comparisons and network meta-analysis, systematic reviews have increased their complexity and pose an even bigger challenge to clinicians trying to understand their results and use them effectively in practice. Also, systematic reviews commonly do not provide a comprehensive picture of all information

that is relevant for decision-making, since patients' values and preferences or resource consideration is not part of their scope.

In recent years, with the consolidation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach,[11] the development of improved standards to judge the trustworthiness of guideline recommendations[12] and the adoption of more strict policies to limit the influence of conflict of interests,[13] clinical practice guidelines have become an alternative to systematic reviews. As systematic reviews, trustworthy guideline recommendations include an updated summary of the evidence about the effect of the interventions, but also provide information regarding patients' values and preferences and resource utilization. Additionally, they feature an actionable recommendation for practice that differentiate between courses of actions that must be offered to (or avoided in) all or almost all the patients (called strong recommendations in the GRADE framework) from decisions that should be individualized to the specific clinical circumstances and patients' preferences (weak recommendations in the GRADE framework). Clinicians, who typically have limited time to review and integrate the evidence, prefer having this guidance accompanying a summary of the evidence.[14]

To effectively use guidelines in practice, however, clinicians need be able to understand and critically assess recommendations. In the second chapter of this thesis, "How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses", we provide a guide to critically assess and use guideline recommendations. Through several examples from real guidelines, we emphasize the factors that make a recommendation trustworthy and illustrate how to use a recommendation in everyday practice according to its strength.

Given that the strength of the recommendation is essential to its interpretation, in the third chapter, "Assessing the Strength of Recommendations: The GRADE Approach", we expand the second chapter with a description of the process used in the GRADE approach to move from the evidence to the recommendations. Also, through practical and real examples, we explain the judgments that guideline panelists have to make during the development process.

The fourth and fifth chapters of this thesis address a critical issue of the guideline development process: the management of the conflict of interests. The development of recommendations requires several judgments from guideline panelists: judgments about the quality of the evidence, about the balance between benefits and harms, about how variable patients' values and preferences are and about to what extent the incremental cost of the intervention is justified by its potential benefits. All of those judgments can be potentially influenced by competing interest of guideline panelists. Indeed, the high prevalence of financial ties of guideline panelists with pharmaceutical industry[15,16] and the potential influence of intellectual conflict of interests[17] and professional loyalties[18] raise concerns regarding the trustworthiness of guideline recommendations and threatens the guideline enterprise as a whole.

In the context of the development of the American College of Chest Physicians 9th edition of the Antithrombotic Guidelines, the executive committee developed and implemented a new strategy for limiting the influence of conflict of interests while allowing the participation of conflicted panelist in some aspect of the recommendation development process. According this strategy, methodologists free from conflict of interests bore primary responsibility for the recommendations. Content area experts and

other panelists with, potentially, both financial and intellectual conflict of interests had input into preparing, summarizing, and interpreting the evidence. However, the intention was to exclude them from the deliberations that ultimately determined the direction and strength of recommendations on which they had conflicts.[13]

A qualitative study conducted before the implementation of this strategy found that methodologists and content expert were uneasy regarding the effects of the new policy.[19]

The fourth chapter of this thesis, "Experiences with a novel policy for managing conflicts of interest of guideline developers: A descriptive qualitative study", presents and analyses a series of interviews conducted at the end of the Antithrombotic Guidelines development process. This qualitative study describes the experiences of methodologists and leading content experts with the new strategy for managing conflicts of interest. The fifth chapter of this thesis, "Low anonymous voting compliance with the novel policy for managing conflicts of interest implemented in the ninth version of the American College of Chest Physicians antithrombotic guidelines", quantitatively assesses the extent to which the new strategy for managing conflicts of interest was respected when panelists had to vote anonymously.

References

1. McColl A, Smith H, White P, Field J. General practitioner's perceptions of the route to evidence based medicine: a questionnaire survey. BMJ. 1998 Jan 31;316(7128):361-5.
2. O'Donnell CA. Attitudes and knowledge of primary care professionals towards evidence-based practice: a postal survey. J Eval Clin Pract. 2004 May;10(2):197-205.
3. Heiwe S, Kajermo KN, Tyni-Lenné R, Guidetti S, Samuelsson M, Andersson IL, Wengström Y. Evidence-based practice: attitudes, knowledge and behaviour among allied health care professionals. Int J Qual Health Care. 2011 Apr;23(2):198-209.
4. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. N Engl J Med 2003; 348:2635-45.
5. Haynes B, Haines A. Barriers and bridges to evidence based clinical practice. BMJ. 1998 Jul 25;317(7153):273-6.
6. Glenton C, Underland V, Kho M, Pennick V, Oxman AD. Summaries of findings, descriptions of interventions, and information about adverse effects would make reviews more informative. J Clin Epidemiol. 2006 Aug;59(8):770-8.
7. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. J Clin Epidemiol. 2010 Jun;63(6):620-6.

8. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med. 1997 Mar 1;126(5):376-80.
9. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010 Sep 21;7(9):e1000326.
10. Hoffmann T, Erueti C, Thorning S, Glasziou P. The scatter of research: cross sectional comparison of randomised trials and systematic reviews across specialties. BMJ. 2012 May 17;344:e3223
11. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64(4):380-2.
12. Institute of Medicine. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, eds. Clinical practice guidelines we can trust. Washington (DC): National Academies Press; 2011. 2p.
13. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010;152(11):738-41
14. Neumann et al. Do clinicians want recommendation. A randomized trial comparing evidence summaries with and without recommendations. Unpublished.
15. Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: A systematic review. PLoS One. 2011;6(10):e25153. Epub 2011 Oct 19.

16. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011 Oct 11;343:d5621

17. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

18. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012 Jul;65(7):725-33.

19. Akl EA, Karl R, Guyatt GH. Methodologists and context experts disagreed regarding managing conflicts of interest of clinical practice guidelines panels. J Clin Epidemiol. 2012 Jul;65(7):734-9.

Chapter 2: How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses

Ignacio Neumann, Elie A. Akl, Per Olav Vandvik, Thomas Agoritsas, Pablo Alonso-Coello, David M. Rind, Nancy Santesso, Paul Elias Alexander, Reem Mustafa, Kameshwar Prasad, Shannon Bates, Holger J. Schunemann, and Gordon H. Guyatt

*This paper has been accepted for publication and will be included in the upcoming third edition of the Journal of American Medical Association "Users' guide to medical literature".

Clinical Scenario:

You are an obstetrician seeing a 31-year-old pregnant woman who had an unprovoked deep venous thrombosis of the leg 5 years ago that was treated with warfarin for 6 months without complication. She is no longer using antithrombotic medication and is otherwise healthy. Given a possible increased risk of thrombosis with pregnancy, you are considering discussing the possibility of low-molecular-weight heparin (LMWH) prophylaxis for the rest of the pregnancy.

To inform your discussion, you search first for an evidence-based recommendation and find the following recommendation from a practice guideline[1]: "For pregnant women at moderate to high risk of recurrent venous thromboembolism (VTE) (single unprovoked VTE, pregnancy- or estrogen-related VTE, or multiple prior unprovoked VTE not receiving long-term anticoagulation), we suggest antepartum prophylaxis with prophylactic- or intermediate- dose LMWH rather than clinical vigilance or routine care (weak recommendation, based on low confidence in effect estimates)."

The statement "weak recommendation, based on low confidence in effect estimates" leaves you uncomfortable. You decide to read further to understand the recommendation and its rationale.

Developing Recommendations

In general, patient management recommendations are developed in the context of clinical practice guidelines. However, you also may find guidance originating from a decision analysis. Similar criteria of credibility apply to both approaches.[2-5]

Practice Guidelines

Practice guidelines are statements that include recommendations intended to optimize patient care. They are, ideally, informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options.[2] To make a recommendation, guideline panelists must define clinical questions, select the relevant outcome variables, retrieve and synthesize all the relevant evidence, rate the confidence in the effect estimates, and, relying on a systematic approach but ultimately also on consensus, move from evidence to recommendations.[6] To fully inform their audience, guideline panels should provide not only their recommendations but also the key information on which their recommendations are based.

Decision Analysis

Decision analysis is a formal method that integrates the evidence regarding the beneficial and harmful effects of treatment options with the values or preferences associated with those effects. Clinical decision analyses are built as structured approaches (decision trees), and authors will usually include 1 or more diagrams showing the structure of the decision trees used for the analysis.

Figure 26-1 shows a simplified decision tree for the scenario of the pregnant woman considering thromboprophylaxis. The patient has 2 options: to use or not use prophylaxis with LMWH. The decision is represented by a square, termed as “decision node.” The lines emanating from the decision node represent the clinical strategies under consideration.

Circles, called "chance nodes," symbolize the different events that can occur following each clinical strategy. Patients may or may not develop a thrombotic or a bleeding event, and the decision analysis requires estimates of the probability of both events. Triangles or rectangles identify outcome states.

The decision analysis also addresses the extent to which each of the outcome events is desirable (no bleeding or thrombotic event) or undesirable (either adverse event) (in technical language, the utility). The combination of the probabilities and utilities allows the decision analyst to determine the relative value of each management option.

The process of decision analysis makes fully explicit all of the elements of the decision so that they are open for debate and modification.[7] When a decision analysis includes costs among the outcomes, it becomes an economic analysis and summarizes trade-offs between health changes and resource expenditure.

Example of a Decision Tree

Returning to Figure 26-1, each arm of the decision (no prophylaxis vs LMWH) has 1 chance node at which 4 possible outcomes could occur (the 4 possible combinations arising from bleeding or not bleeding and from having a thrombosis or not having a thrombosis). The figure depicts the probabilities associated with the decision. In the no-prophylaxis strategy, patients would have a probability of bleeding and having a thrombosis of 0.1%; a probability of bleeding and not having a thrombosis of 1.3%; a probability of not bleeding but having a thrombosis of 8%; and a probability of not bleeding and not having a thrombosis of 90.6%. With the LMWH prophylaxis strategy, the probability of bleeding and having a thrombosis is 0.06%; the probability of bleeding

and not having a thrombosis is 2%; the probability of not bleeding but having a thrombosis is 2.9%; and the probability of not bleeding and not having a thrombosis is 95%.[1]

The figure also presents the values associated with each health state on a scale from 0 to 1, with 1 representing the utility of full health and 0 representing the utility of death. In the no-prophylaxis strategy, the health state without any negative outcome (no thrombosis or bleeding) represents full health, a utility of 1.0. The occurrence of a thrombosis or bleeding event decreases the value of the health state to 0.45 in the case of thrombosis and to 0.38 in the case of bleeding. When both negative outcomes occur at the same time, the corresponding utility is even lower: 0.25. In the LMWH arm, the addition of the burden of treatment slightly decreases the utility of the 4 health states.

The final step in the decision analysis is to calculate the total expected value—the sum of the probabilities and utilities associated with each outcome—for each possible course of action. Given the particular set of probabilities and utilities we have presented, the estimated value of the no-prophylaxis branch would be $(0.906 \times 1.0) + (0.080 \times 0.45) + (0.013 \times 0.38) + (0.001 \times 0.25)$, which is 0.947. The value of the LMWH branch would be $(0.950 \times 0.98) + (0.029 \times 0.43) + (0.020 \times 0.36) + (0.0006 \times 0.24)$, which is 0.950. In this example, the prophylaxis strategy is more desirable, but the difference in the expected values between the 2 options—called "relative utility"—is relatively small.

The model presented in Figure 26-1 is oversimplified in a number of ways. For example, it does not take into account the possibility of fatal events or potential long-term morbidity (for example, after an intracranial bleeding, or the development of post-thrombotic syndrome). Also, it does not consider the time in the health states. For

instance, having a major bleeding without any complication may appreciably reduce the utility during the episode, but almost all patients will return to a perfect health state relatively quickly. Multistate transition models using simulation—termed Markov models—permit analyses that are closer to real life. For example, an analysis using multistate transition models concluded that for patients like the one presented in the opening scenario and in the decision tree (high risk for VTE recurrence), antepartum prophylaxis with LMWH is a cost-effective use of resources.[9]

Assessing Recommendations

Box 26-1 presents our guidance for determining the extent to which a guideline or decision analysis will provide trustworthy recommendations.

Is the Clinical Question Clear and Comprehensive?

The most useful patient management recommendations from guidelines and decision analyses will use a standardized format detailing precisely the recommended actions, the alternatives with which they are compared, to whom they apply, and under what circumstances.

Is the Recommended Intervention Clear and Actionable?

Sometimes, recommendations are too vague to be helpful. Consider, for instance, this recommendation from a clinical practice guideline:[10] "For both outpatients and inpatients with diabetic foot infection, clinicians should attempt to provide a well-coordinated approach by those with expertise in a variety of specialties, preferably by a

multidisciplinary diabetic foot care team." What remains unclear in this recommendation is the level of obligation in the "attempt," what is involved in making care "well-coordinated," and which specialties are included in the "variety."

In contrast, another guideline from the National Foundation for Health Care Excellence[11] makes clear what is being recommended: "We recommend that a multidisciplinary foot care team manage the care of patients with diabetic foot problems who require inpatient care. The multidisciplinary foot care team should include a diabetologist, a surgeon with the relevant expertise, a diabetes nurse specialist, a podiatrist and a tissue viability nurse."

Is the Alternative Clear?

When guideline panelists develop recommendations, they choose a specific course of action over others. If the alternative is not clear, the significance of the recommendation will remain obscure. For example, in the recommendation "Uterine massage is recommended for the treatment of postpartum hemorrhage,"[12] the absence of an explicit alternative may introduce challenges in the interpretation. Are the panelists suggesting performing uterine massage as a first-line treatment in preference to other therapeutic measures, or are they recommending it in addition to other concomitant measures? By comparing the recommendation with others within the guideline, it is possible to infer that panelists meant that uterine massage should be used in addition to other measures, and not as a single intervention, but recommendation statements should be clear enough to be interpreted without having to read the full guideline. In contrast, the recommendation "We recommend isotonic crystalloids...in preference to...colloids for the

initial intravenous fluid resuscitation of women with postpartum hemorrhage"[12] offers a clearer message by making the alternative explicit.

As you may have noticed, in both recommendations regarding the management of diabetic foot problems presented in the previous section, the control group is not clearly defined. Although the option of "no foot care team" seems to be the implicit comparator, it is not clear what this management strategy entails.

Clinicians using a decision analysis will not face the problem of ambiguous alternatives, since the options in comparison are explicit.

Were All the Relevant Outcomes Important to Patients Explicitly Considered?

The balance between the benefits and the harms of the interventions will depend on what outcomes are considered. Clinicians should judge whether or not the guideline panel, or the decision analysts, included all patient-important outcomes.

For example, the 8th edition of the Antithrombotic Guidelines of the American College of Chest Physicians (ACCP) recommended the use of elastic stockings for patients with stroke who have contraindications to anticoagulants.[13] The 9th edition of the Antithrombotic Guidelines (AT9) suggested against its use.[14] Both guideline panels considered the outcomes of mortality, pulmonary embolism, and symptomatic deep venous thrombosis, but AT9 panelists also considered that elastic stockings produce a 4-fold increase in the risk of skin complications: 39 more per 1000 patients treated for 1 month (95% CI, 17-77 more per 1000).[15] The additional consideration of skin complications is responsible for the change in recommendations.

Outcomes typically considered as patient-important include mortality, morbidity (eg, major bleeding, acute exacerbation of a chronic disease, hospital admission), and patient-reported outcomes (eg, quality of life, functional status). Surrogate outcomes (eg, lipid levels, bone density, cognitive function tests) are variably associated with patient-important outcomes, but are never important in and of themselves.

In addition, the 8th edition of the Antithrombotic Guidelines of the ACCP suggested international normalized ratio (INR) monitoring at an interval of no longer than every 4 weeks in patients treated with vitamin K antagonists.[16] This recommendation was primarily based on studies showing that frequent monitoring increased the time in therapeutic INR range—a surrogate outcome. The 9th edition of the Antithrombotic Guidelines, however, suggested an INR testing interval of up to 12 weeks rather than every 4 weeks.[17] This recommendation was based on studies showing no increase in thrombotic events or major bleeding with monitoring every 12 weeks. Both recommendations were based on explicitly defined outcomes. However, the outcomes were surrogate in the first case and—more appropriately—patient important in the second.

Outcomes not plausibly influenced by the intervention are typically not relevant for decision-making, and therefore may not be considered. For example, mortality is a very important outcome; however, it is not relevant for the decision of whether or not use intranasal antihistamines for the treatment of allergic rhinitis, since the intervention does not plausibly affect the probability of dying.

Were the Recommendations Based on the Current Best Evidence?

Guideline panelists and decision analysts should base their estimates of the benefits and harms of the intervention and their evaluation of the associated confidence in effect estimates on current or updated systematic reviews, preferably those that include meta-analysis. In the absence of such meta-analytic systematic reviews, guideline panelists may conduct their own reviews or provide less-systematic evidence summaries. Clinicians should look for a description of the process used to identify and summarize the relevant evidence and should judge to what extent this process is credible. Clinicians also should check the date on which the literature search was conducted.

Recommendations that do not use the best current evidence risk promoting suboptimal or even harmful care. For example, for several years guideline panels ignored a substantial body of evidence suggesting the effectiveness of prophylaxis with quinolones in patients with postchemotherapy neutropenia.[18] Only in its 2010 guidelines did the Infectious Diseases Society of America suggest the prophylactic use of antibiotics in this population.[19] This highlights the necessity for rapid and sometimes frequent updating of guidelines in areas under active investigation.

Are Values and Preferences Appropriately Specified for Each Outcome?

Assessing treatment effects on outcomes is largely a question of measurement and a matter of science. Assigning preferences to outcomes is a matter of values. Consider, for example, the outcomes associated with routine mammographic screening in women aged 40 to 49 years: there is a very small and questionable reduction of breast cancer mortality and a relatively high probability of a false-positive (which typically leads to unnecessary

follow-up testing, and sometimes to unnecessary biopsy of the breast).[20] A guideline panel must consider the value attached to each of these 2 outcomes when trading them off to develop a recommendation. A panel assigning a higher value to the very small reduction in cancer mortality would support the screening, while a panel assigning a higher value to avoiding unnecessary procedures would not. Consequently, clinicians should look for explicit statements regarding the values and preferences used to inform the recommendation.

Whose values should drive recommendations? Under ideal circumstances, recommendations should be based on a systematic review of relevant studies exploring patients' values and preferences;[21] unfortunately, such evidence is still rare. In the absence of a body of empirical evidence about patients' values and preferences, guideline panels or decision analysts may fall back on the experience of clinicians who regularly engage in shared decision making. Another alternative is the involvement of representative patients and consumers in the recommendation development process.[22] However, ensuring that those involved—clinicians or patients—will be able to represent typical patients is challenging and perhaps only partly achievable.

Whatever the source of values and preferences, it is possible to make them explicit and transparent. Unfortunately, failure to do so remains the most common serious deficit in current practice guidelines. In contrast, decision analysis requires explicit and quantitative specification of values, as each outcome is assigned a given health utility. However, although the values and preferences in a decision analysis may be explicit, their source may be problematic. For example, a systematic review of 54 cost-utility analyses (including 45 decisions analyses) in child health found that the source used for valuing

health states was the authors' own judgment in 35% of the analyses, and in another 11% the source of values and preferences was not stated.[23]

Do the Authors Indicate the Strength of Their Recommendations?

Trustworthy recommendations should specify the strength of the recommendations, and also a rating of the confidence in effect estimates that support the recommendations (also known as quality of evidence).[2] Sensitivity analyses are used to explore the strength of the conclusions arising from a decision analysis.

Grades of Recommendation

There are dozens of grading systems for recommendations.[24] However, the 3 most commonly used approaches are GRADE[25] and those used by the American Heart Association (AHA)[26] and the US Preventive Services Task Force (USPSTF).[27] A detailed discussion of the differences between these systems is beyond the scope of this chapter; we will, however, mention 2 important similarities.

The 3 systems feature a rating for confidence in effect estimates (ie, quality of evidence). Confidence in the effect estimates represents the extent to which the estimates are sufficiently credible to support a particular recommendation. The GRADE approach specifies 4 levels of confidence: high, moderate, low, and very low. The AHA and USPSTF systems specify 3 levels of confidence: A, B, and C in AHA, and high, moderate, and low in USPSTF.

The 3 systems share another critical feature: they differentiate between recommendations that should be applied (or avoided) in all, or almost all, patients (ie, strong

recommendations) from those which require individualization to the patient's values, preferences, and circumstances (ie, weak recommendations) (Figure 26-2).

Sensitivity Analysis

Decision analysts use sensitivity analyses, the systematic exploration of the uncertainty in the data, to vary estimates for downsides, benefits, and values and to determine the impact of these varying estimates on expected outcomes. Sensitivity analysis asks the question: to what extent is the relative utility of the alternatives affected by the uncertainties in the estimates of the likelihood or value of the outcomes? To the extent that the result of the decision analysis does not change with varying probability estimates and varying values, clinicians can consider the recommendation a strong one. When the final decision shifts with different plausible values of probabilities or values, the conclusion becomes much weaker: the right choice may differ given the true probabilities, and patients' choices are likely to vary according to their preferences.

Is the Evidence Supporting the Recommendations Easily Understood?

For Strong Recommendations, Is the Strength Appropriate?

The message to the clinician from strong recommendations is "just do it." Recommendations that are inappropriately graded as strong may therefore have substantial undesirable consequences.

High confidence in the effect estimates will support a strong recommendation if the desirable consequences considerably outweigh the undesirable consequences, if there is reasonable confidence and limited variability in patients' values and preferences, and if

the benefits of the proposed course of action justify its cost. When there is substantial uncertainty regarding the effects of the intervention (low confidence in the effect estimates), clinicians should generally expect weak recommendations.

Sometimes, guideline panels can appropriately offer strong recommendations despite low or very-low confidence in effect estimates. Table 26-1 presents 5 paradigmatic situations in which this can occur. Clinicians should carefully examine a strong recommendation based on low or very-low confidence. If it does not correspond to any of the situations described in Table 26-1, it is likely that the recommendation was inappropriately graded.

For example, a systematic survey of the Endocrine Society guidelines between 2005 and 2011 found that 121 of the total of 357 recommendations identified were strong recommendations based on low or very-low confidence in effect estimates. Of these 121, only 35 (29%) were consistent with one of the situations presented in Table 26-1, and thus clearly appropriate.[28] This result highlights the need for caution when facing strong recommendations based on low or very-low confidence in effect estimates.

In decision analysis, the parallel to strong recommendations occur when the relative utility of the management options changes little, and the preferred alternative does not change, after varying probability estimates and varying values. Clinicians should look for a table that lists which variables were included in their sensitivity analyses, what range of values they used for each variable, and which variables, if any, altered the relative desirability of the management strategies under consideration.

Ideally, decision analysts will subject all of their probability estimates to a sensitivity analysis. The range over which they will test should depend on the source of the data. If the estimates come from large randomized trials with low risk of bias and narrow CIs, the

range of estimates tested can be narrow. When risk of bias is greater, or estimates of benefits and downsides less precise, sensitivity analyses testing a wide range of values become appropriate. Decision analysts also should test utility values with sensitivity analyses, with the range of values again determined by the source of the data. If large numbers of patients or knowledgeable and representative members of the general public gave similar ratings to the outcome states, investigators can use a narrow range of utility values in the sensitivity analyses. If the ratings came from a small group of raters, or if the individuals provided widely varying estimates of typical utilities, then investigators should use a wider range of utility values in the sensitivity analyses.

For Weak Recommendations, Does the Information Facilitate Shared Decision Making?

Recommendations—in particular, weak recommendations—should explicitly provide the key underlying information necessary to act on the recommendation. In guidelines, this information is typically found in the remarks section, in the recommendation rationale, or in tables accompanying the recommendation. The GRADE Working Group, in collaboration with the Cochrane Collaboration, has designed a specific table for this purpose: the Summary-of-Findings (SoF) Table. This table provides the confidence ratings for all the important outcomes and the associated estimates of relative and absolute effects. Table 26-2 shows a SoF table relevant for the clinical scenario presented at the beginning of this chapter. As we will discuss later, SoF tables can facilitate shared decision making.[29] The absolute measures of effect you will find in GRADE SoF tables are typically presented within the decisions trees in decision analyses.

Was the Influence of Conflict of Interests Minimized?

The judgments involved in the interpretation of the evidence and deciding on the final recommendation may be vulnerable to conflicts of interest. In medicine, guideline panelists frequently—and decision-analysis authors sometimes—report financial ties with the pharmaceutical industry.[30-32] Nonfinancial conflicts of interests are also common and may have even greater effect than financial conflicts.[33,34] These include intellectual conflicts (eg, previous publication of studies relevant to a recommendation) and professional conflicts (eg, radiologists making recommendations about breast cancer screening or urologists recommending prostate cancer screening).[35,36]

Clinicians can check the conflict of interest statements of the guideline panelists or decision analysts, usually found at the beginning or end of a publication or in a supplementary file. Just as important, clinicians should check what strategies were implemented to manage these conflicts of interest. Guidelines or decision analyses with a large representation of panelists without conflicts of interest, that have placed nonconflicted participants in positions of authority, or that have implemented rules to limit the influence of both financial and nonfinancial conflicts of interest are more credible than those that have not. Guidelines that excluded conflicted experts are likely to have limited the influence of conflicts of interest but may have compromised the credibility of the guidelines and possibly threatened their acceptability. Clinicians also can check whether recommendations were collected and managed for the whole guideline or on a recommendation-by-recommendation basis. The influence of potential conflicts of interest may be diminished with the latter approach.

The AT9 guidelines provide an example of implementation of a number of these strategies.[34] A nonconflicted methodologist was chosen as the chair of each of the 14 panels making recommendations and was primarily responsible for that chapter. The chair and 2 other members of the executive committee ultimately responsible for the whole guideline were nonconflicted methodologists. Both financial and intellectual conflicts of interest were assessed on a recommendation-by-recommendation basis. Panelists with major conflicts were in principle excluded from participation in decision-making. Challenges in implementing this approach highlight the efforts required to arrive at an optimal strategy for managing conflict of interest.[37,38]

Using the Guide

Is the Clinical Question Clear and Comprehensive?

The recommendation presented at the beginning of this chapter clearly specifies what is being proposed ("antepartum prophylaxis with prophylactic- or intermediate- dose LMWH") and what was the comparison ("rather than clinical vigilance or routine care").[1] As we can see in Table 26-2, guideline panelists considered the outcomes of symptomatic thromboembolism, major bleeding, and burden of treatment—the outcomes likely important to patients.

Was the Recommendation Based on the Best Current Evidence?

In the methods section of the published AT9 guidelines, we find the following description: "To identify the relevant evidence, a team…conducted literature searches of Medline, the Cochrane Library, and the Database of Abstracts of Reviews of Effects…for

systematic reviews and another for original studies," and "The quality of reviews was assessed…and wherever possible, current high-quality systematic reviews were used as the source of summary estimates."[39] This strategy ensured that estimates were based on best current evidence at the time the recommendation was issued.

Are Values and Preferences Associated With Outcomes Appropriately Specified?

Guideline authors noted that a systematic review of patient preferences for antithrombotic treatment did not identify any studies of pregnant women. A rating exercise of different outcomes among experienced clinicians participating on the guideline suggested that 1 episode of VTE (deep venous thrombosis or pulmonary embolism) is more or less equivalent to 1 major extracranial bleed. Panelists' clinical experience suggested that the majority of women, but not all, would choose long-term prophylaxis when confronted with the burden of self-injecting with LMWH over several months, suggesting a relatively high value on preventing VTE and a relatively high tolerance for self-injection. These values and preferences were used to develop the recommendation.

Do the Authors Indicate the Strength of Their Recommendations?

The recommendation was classified as "weak" using the GRADE approach.

Is the Evidence Supporting the Recommendation Easily Understood?

The recommendation was accompanied by a SoF table (Table 26-2) that provides absolute estimates for the outcomes important to patients. We will discuss subsequently how this information can help with shared decision making.

Was the Influence of Conflict of Interests Minimized?

As we described earlier, the AT9 guidelines implemented a number of the strategies to diminish the influence of conflict of interest on recommendations.

How Should You Use Recommendations?

Strong Recommendations

If the panel's assessment is astute, clinicians can apply strong recommendations to all or almost all the patients in all or almost all circumstances without thorough—or even cursory—review of the underlying evidence and without a detailed discussion with the patient. The same is true for decision analysis when the utility of 1 alternative is substantially greater than the other and this relative utility is robust to sensitivity analyses. Whether discussion of the evidence with patients might sometimes still be helpful in such circumstances—for instance, whether it may increase adherence to treatment—remains uncertain.

For example, the ARIA (Allergic Rhinitis and its Impact on Asthma) guideline recommended intranasal glucocorticoids rather than intranasal antihistamines for treatment of allergic rhinitis in adults (strong recommendation).[40] This recommendation was based on an important reduction of symptoms with glucocorticoids (rhinorrhea, nasal blockage and itching) with no important adverse events. The effect estimates came from a systematic review of randomized trials with low risk of bias, consistent results across trials, precise effects (narrow CIs), and results applicable to the population. The guideline panel's inference that all, or almost all, informed patients would choose the

glucocorticoids is eminently reasonable. Therefore, a detailed discussion with the patients about the benefits and potential harms of intranasal glucocorticoids over intranasal antihistamines will not be necessary.

There will always be idiosyncratic circumstances in which clinicians should not adhere to even strong recommendations. For instance, aspirin in the context of myocardial infarction warrants a strong recommendation, but it would be a mistake to administer the treatment to a patient who is allergic to aspirin. Such idiosyncratic situations are, fortunately, unusual.

Weak Recommendations

With careful consideration of the evidence, as well as of patient's values and preferences, many recommendations are weak, even in clinical fields with a large body of randomized trials and systematic reviews. For instance, two-thirds of more than 600 recommendations issued in AT9 were weak.[17]

Because weak recommendations are typically sensitive to patients' values and preferences, a shared decision-making approach involving a discussion with the patient addressing the potential benefits and harms of the proposed course of action is the optimal way to ensure that decisions reflects both the best evidence and patients' values and preference. To use weak recommendations, clinicians need to understand the underlying evidence.

For example, the American College of Physicians suggested the use of cholinesterase inhibitors or memantine in patients with dementia (weak recommendation).[41] This recommendation is based on evidence from randomized trials warranting high confidence in a small benefit of the drugs in slowing the deterioration of cognition and global function. Guideline panelists pointed out that, if quality of life is judged as poor—in particular, with more advanced dementia—family members may not view the limited slowing of dementia progression as a desirable goal. Moreover, the magnitude of the effect is small and there are adverse effects associated with the drugs. The panel then reasonably expected that informed patients (or their families) would make different choices.

Clinical Scenario Resolution

After reviewing this guide, and specifically the information in Table 26-2, you decide that the recommendation is trustworthy and you plan to engage patients like the one presented in the opening scenario in shared decision-making. When you meet with the patient, you start by discussing the benefits of LMWH during pregnancy vs no treatment (51 fewer cases of symptomatic VTE per 1000 women), followed by information about adverse effects (7 more maternal bleeds per 1000 women followed over the pregnancy and postpartum), and you mention the potential burden of treatment that daily injections for several months will represent (low confidence in effect estimates for all outcomes aside from the burden of injections). If the guideline panel is correct, most patients will place a higher value in lowering the risk of a thrombotic event and less on the uncertain small

increase in the risk of bleeding and the certain burden of treatment. Such patients will choose prophylaxis. If the panel is correct, however, some patients will decline therapy. Thus, shared decision making is required to ensure the patient understands the best evidence available and the decision is consistent with the patient's values and preference. You are not surprised when the patient chooses VTE prophylaxis.

Tables and Figures

Box 26-1. Users' Guides for Assessing Treatment Recommendations

Is the clinical question clear and comprehensive?

Is the recommended intervention clear and actionable?

Is the alternative clear?

Were all the relevant outcomes important to patients explicitly considered?

Was the recommendation based on the best current evidence?

Are values and preferences associated with the outcomes appropriately specified?

Do the authors indicate the strength of their recommendations?

Is the evidence supporting the recommendation easily understood?

For strong recommendations, is the strength appropriate?

For weak recommendations, does the information provided facilitate shared decision-making?

Was the influence of conflict of interests minimized?

Table 26-1. Five Paradigmatic Situations That Justify Strong Recommendations Based on Low or Very-Low Confidence

Paradigmatic Situation	Confidence in Effect Estimates for Health Outcomes (Quality of Evidence)		Balance of Benefits and Harms	Values and Preferences	Resource Considerations	Recommendation	Example
	Benefits	Harms					
Life-threatening situation	Low or very low	Immaterial (very low to high)	Intervention may reduce mortality in a life-threatening situation; adverse events not prohibitive	A very high value is placed on an uncertain but potentially life-preserving benefit	Small incremental cost (or resource use) relative to the benefits justify the intervention	Strong recommendation in favor	Indirect evidence from seasonal influenza suggests that patients with avian influenza may benefit from the use of oseltamivir (low confidence in effect estimates). Given the high mortality of the disease and the absence of effective alternatives, the WHO made a strong recommendation in favor of the use of oseltamivir rather than no treatment in patients with avian influenza.[42]
Uncertain benefit, certain harm	Low or very low	High or moderate	Possible but uncertain benefit; substantial established harm	A much higher value is placed on the adverse events in which we are confident than in the benefit, which is uncertain	High incremental cost (or resource use) relative to the benefits may not justify the intervention	Strong recommendation against	In patients with idiopathic pulmonary fibrosis, treatment with azathioprine plus prednisone offers a possible but uncertain benefit in comparison with no treatment. The intervention, however, is associated with a substantial established harm. An international guideline made a recommendation against the combination of corticosteroids plus azathioprine in patients with idiopathic pulmonary fibrosis.[42]
Potential equivalence, one option clearly less risky or costly	Low or very low	High or moderate	Magnitude of benefit apparently similar—though uncertain—for alternatives; we are confident less harm or cost for one of the competing alternatives	A high value is placed on the reduction in harm	High incremental cost (or resource use) relative to the benefits may not justify one of the alternatives	Strong recommendation for less harmful/less expensive	Low-quality evidence suggests that initial *Helicobacter pylori* eradication in patients with early stage extranodal marginal zone (MALT) B-cell lymphoma results in similar rates of complete response in comparison with the alternatives of radiation therapy or gastrectomy, but with high confidence of less harm, morbidity, and cost. Consequently, UpToDate made a strong recommendation in favor of *H pylori* eradication rather than radiotherapy in patients with MALT lymphoma.[44]
High confidence in similar benefits, one option potentially more risky or costly	High or moderate	Low or very low	Established that magnitude of benefit is similar for alternative management strategies; best (though uncertain) estimate is that one alternative has appreciably greater harm	A high value is placed on avoiding the potential increase in harm	High incremental cost (or resource use) relative to the benefits may not justify one of the alternatives	Strong recommendation against the intervention with possible greater harm	In women requiring anticoagulation and planning conception or in pregnancy, high confidence estimates suggest similar effects of different anticoagulants. However, indirect evidence (low confidence in effect estimates) suggests potential harm to the unborn infant with oral direct thrombin (eg, dabigatran) and factor Xa inhibitors (eg, rivaroxaban, apixaban). The AT9 guidelines recommended against the use of such anticoagulants in women planning conception or in pregnancy.[1]
Potential catastrophic harm	Immaterial (very low to high)	Low or very low	Potential important harm of the intervention, magnitude of benefit is variable	A high value is placed on avoiding potential increase in harm	High incremental cost (or resource use) relative to the benefits, may not justify the intervention	Strong recommendation against the intervention	In males with androgen deficiency, testosterone supplementation likely improves quality of life. Low-confidence evidence suggests that testosterone increases cancer spread in patients with prostate cancer. The US Endocrine Society made a recommendation against testosterone supplementation in patients with prostate cancer.[45]

Abbreviations: AT9, Antithrombotic Guidelines, 9th Ed; WHO, World Health Organization

Table 26-2. Summary of Findings (SoF) Table: Antepartum and Postpartum Prevention of Prevention of Venous Thromboembolism (VTE) With Prophylactic Dose of Low-Molecular-Weight Heparin vs No Prophylaxis in Pregnant Women With Prior VTE[1]

Outcome	Relative Effect (95% CI)	Anticipated Absolute Effects During Pregnancy		Confidence in the Estimates of the Effect
		Risk Without Prophylaxis	Risk Difference With LMWH	
Symptomatic VTE	RR, 0.36 (0.20-0.67)	**Low Risk**		**Low** due to indirectness[b] and imprecision[c]
		20 VTE per 1000	13 fewer VTE per 1000 (from 16 to 7 fewer)	
		Intermediate and High Risk[a]		
		80 VTE per 1000	51 fewer VTE per 1000 (from 65 to 30 fewer)	
Major bleeding	RR, 1.57 (1.32-1.87)[d]	**Antepartum Period**		**Low** due to indirectness[a] and imprecision[f]
		3 bleeds per 1000	1 more bleed per 1000 (from 1 to 3 more)[e]	
		Postpartum Period		
		10 bleeds per 1000	6 more bleeds per 1000 (from 3 to 8 more)[d]	
Burden of treatment	---	No incremental burden	Daily injections	**High**

Abbreviations: RR, risk ratio; VTE, venous thromboembolism; LMWH: low-molecular-weight heparin.
[a]Single unprovoked VTE, pregnancy-related or estrogen-related VTE, or multiple prior unprovoked VTE not receiving long-term anticoagulation.
[b]Population is indirect (ie, did not include pregnant women).
[c]95% CI includes marginal benefit.
[d]Relative effect estimate based on the systematic review by Collins et al.[46]
[e]Absolute risk estimates for major bleeding in women using LMWH based on the systematic review by Greer et al.[47]
[f]95% CI includes marginal harm.

Adapted from Bates et al.[1]

Figure 26-1. Diagram of a Simplified Decision Tree

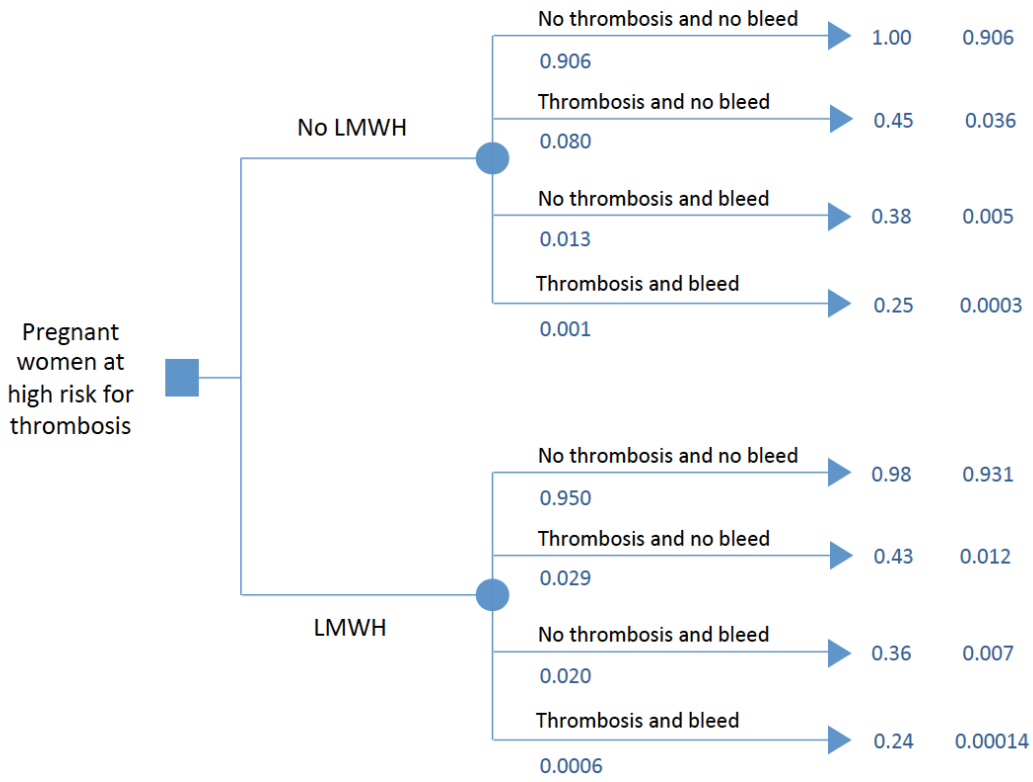
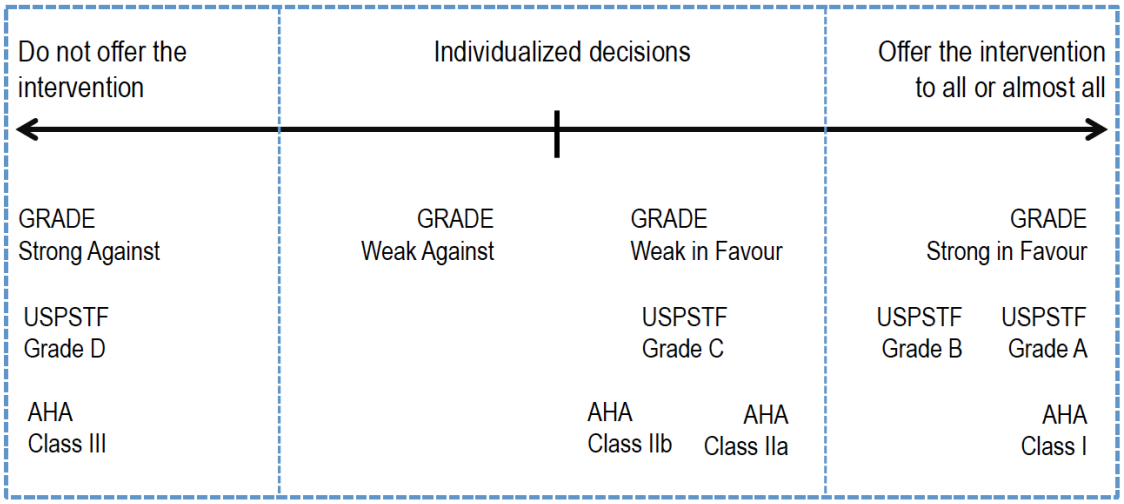


Figure 26-2. Direction and Strength of Recommendations in Different Grading Systems



References

1. Bates SM, Greer IA, Middeldorp S, Veenstra DL, Prabulos AM, Vandvik PO. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e691S-736S.
2. Institute of Medicine. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, eds. Clinical practice guidelines we can trust. Washington (DC): National Academies Press; 2011. 2p.
3. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. Annals of Internal Medicine. 2011;154(11):774-5.
4. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. Annals of Internal Medicine. 2012;156(7):525-31.
5. Shekelle P, Woolf S, Grimshaw JM, Schünemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. Implement Sci. 2012;7:62.
6. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2013 Dec 16. [Epub ahead of print]

7. Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: a progress report. Ann Intern Med. 1987;106(2):275-291
8. Jackson R, Lawes CM, Bennett DA, Milne RJ, Rodgers A. Treatment with drugs to lower blood pressure and blood cholesterol based on an individual's absolute cardiovascular risk. Lancet. 2005;365(9457):434-441.
9. Johnston JA, Brill-Edwards P, Ginsberg JS, Pauker SG, Eckman MH. Cost-effectiveness of prophylactic low molecular weight heparin in pregnant women with a prior history of venous thromboembolism. Am J Med. 2005;118(5):503-14.
10. Lipsky BA, Berendt AR, Cornia PB, Pile JC, Peters EJ, Armstrong DG, Deery HG, Embil JM, Joseph WS, Karchmer AW, Pinzur MS, Senneville E; Infectious Diseases Society of America. 2012 Infectious Diseases Society of America clinical practice guideline for the diagnosis and treatment of diabetic foot infections. Clin Infect Dis. 2012;54(12):e132-7.
11. National Institute for Health and Care Excellence. Diabetic foot problems: Inpatient management of diabetic foot problems (CG119). London: National Institute for Health and Care Excellence. 2011
12. WHO recommendations for the prevention and treatment of postpartum haemorrhage. Geneva, World Health Organization (WHO), 2012.
13. Albers GW, Amarenco P, Easton J, Sacco RL, Teal P. Antithrombotic And Thrombolytic Therapy For Ischemic Stroke: American College Of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008;133(6_suppl):630S-669S.

14. Lansberg MG, O'Donnell MJ, Khatri P, et al. Antithrombotic and thrombolytic therapy for ischemic stroke: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e601S-36S.
15. Dennis M, Sandercock PA, Reid J, et al; CLOTS Trials Collaboration. Effectiveness of thigh-length graduated compression stockings to reduce the risk of deep vein thrombosis after stroke (CLOTS trial 1): a multicentre, randomised controlled trial. Lancet. 2009;373(9679):1958-1965.
16. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G; American College of Chest Physicians. Pharmacology and management of the vitamin K antagonists: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008;133(6 Suppl):160S-198S.
17. Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, Kovacs MJ, Svensson PJ, Veenstra DL, Crowther M, Guyatt GH; American College of Chest Physicians. Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e152S-84S.
18. Hughes WT, Armstrong D, Bodey GP, et al. 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. Clin Infect Dis. 2002;34(6):730-51.
19. Freifeld AG, Bow EJ, Sepkowitz KA, Boeckh MJ, Ito JI, Mullen CA, Raad II, Rolston KV, Young JA, Wingard JR; Infectious Diseases Society of America.

Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the Infectious Diseases Society of America. Clin Infect Dis. 2011;52(4):e56-93.

20. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009;151(10):716-26, W-236.
21. MacLean S, Mulla S, Akl EA, Jankowski M, Vandvik PO, Ebrahim S, McLeod S, Bhatnagar N, Guyatt GH; American College of Chest Physicians. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):e1S-23S.
22. Nilsen ES, Myrhaug HT, Johansen M, Oliver S, Oxman AD. Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. Cochrane Database Syst Rev. 2006;(3):CD004563.
23. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. Pediatrics. 2005;115(5):e600-14.
24. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. BMC Health Serv Res. 2004;4(1):38.

25. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64(4):380-2.

26. American College of Cardiology Foundation and American Heart Association. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines (2010). http://my.americanheart.org/professional/StatementsGuidelines/PoliciesDevelopment/Development/Methodologies-and-Policies-from-the-ACCAHA-Task-Force-on-Practice-Guidelines_UCM_320470_Article.jsp. Accessed January 13, 2014.

27. US Preventive Services Task Force. Grade Definitions. http://www.uspreventiveservicestaskforce.org/uspstf/grades.htm. Accessed January 13, 2014.

28. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. J Clin Endocrinol Metab. 2013;98(8):3246-52.

29. Treweek S, Oxman AD, Alderson P, Bossuyt PM, Brandt L, Bro¿ek J, Davoli M, Flottorp S, Harbour R, Hill S, Liberati A, Liira H, Schünemann HJ, Rosenbaum S, Thornton J, Vandvik PO, Alonso-Coello P; DECIDE Consortium. Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): protocol and preliminary results. Implement Sci. 2013 Jan 9;8:6.

30. Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. PLoS One. 2011;6(10):e25153.

31. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011;343:d5621.

32. Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA. 2002;287(5):612-7.

33. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2(8):e124. Epub 2005 Aug 30.

34. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010;152(11):738-41

35. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012;65(7):725-33.

36. Dahm P, Kunz R, Schünemann H. Evidence-based clinical practice guidelines for prostate cancer: the need for a unified approach. Curr Opin Urol. 2007;17(3):200-7.

37. Neumann I, Karl R, Rajpal A, Akl EA, Guyatt GH. Experiences with a novel policy for managing conflicts of interest of guideline developers: a descriptive qualitative study. Chest. 2013;144(2):398-404.

38. Neumann I, Akl EA, Valdes M, Bravo S, Araos S, Kairouz V, Schünemann H, Guyatt GH. Low Anonymous Voting Compliance With the Novel Policy for

Managing Conflicts of Interest Implemented in the 9th Version of the American College of Chest Physicians Antithrombotic Guidelines. Chest. 2013;144(4):1111-1116.

39. Guyatt GH, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, Crowther M, Vandvik PO, Eikelboom JW, McDonagh MS, Lewis SZ, Gutterman DD, Cook DJ, Schünemann HJ. Methodology for the development of antithrombotic therapy and prevention of thrombosis guidelines: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141(2 Suppl):53S-70S.
40. Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, van Wijk RG, Ohta K, Zuberbier T, Schünemann HJ; Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision.J Allergy Clin Immunol. 2010 Sep;126(3):466-76.
41. Qaseem A, Snow V, Cross JT Jr, Forciea MA, Hopkins R Jr, Shekelle P, Adelman A, Mehr D, Schellhase K, Campos-Outcalt D, Santaguida P, Owens DK; American College of Physicians/American Academy of Family Physicians Panel on Dementia. Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. Ann Intern Med. 2008;148(5):370-8.
42. Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, Yazdanpanah Y, Beigel J, Chotpitayasunondh T, Del Mar C, Farrar J, Tran TH, Ozbay B, Sugaya N, Fukuda K, Shindo N, Stockman L, Vist GE, Croisier A, Nagjdaliyev A, Roth C, Thomson G, Zucker H, Oxman AD; WHO Rapid Advice

Guideline Panel on Avian Influenza. WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. Lancet Infect Dis. 2007;7(1):21-31.

43. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE Jr, Kondoh Y, Myers J, Müller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schünemann HJ; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824.
44. Freedman AS, Lister A, Connor RF. Management of gastrointestinal lymphomas. In: UpToDate, Basow, DS (Ed), UpToDate, Waltham, MA; 2013.
45. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010;95(6):2536-59.
46. Collins R, Scrimgeour A, Yusuf S, Peto R. Reduction in fatal pulmonary embolism and venous thrombosis by perioperative administration of subcutaneous heparin. Overview of results of randomized trials in general, orthopedic, and urologic surgery. N Engl J Med. 1988;318(18):1162-73.

47. Greer IA, Nelson-Piercy C. Low-molecular-weight heparins for thromboprophylaxis and treatment of venous thromboembolism in pregnancy: a systematic review of safety and efficacy. Blood. 2005;106(2):401-7.

Authors' contributions:

Dr. Neumann contributed to the conception of the chapter, data collection and drafting of the article. All the coauthors contributed to the conception of the chapter and drafting of the article

Chapter 3: Assessing the Strength of Recommendations: The GRADE Approach

Ignacio Neumann, Elie A. Akl, Per Olav Vandvik, Pablo Alonso-Coello, Nancy Santesso, M. Hassan Murad, Frederick Spencer, Holger J. Schunemann, and Gordon H. Guyatt

*This paper has been accepted for publication and will be included in the upcoming third edition of the Journal of American Medical Association "Users' guide to medical literature".

Clinical Scenario:

You are a primary care practitioner considering the possibility of the use of aspirin for primary prevention of serious cardiovascular events and cancer in a 60-year-old man. He has hypertension, adequately controlled with thiazides, but he is otherwise healthy; he does not have diabetes or dyslipidemia, does not smoke, and has no family history of heart disease.

To inform your decision, you search first for an evidence-based recommendation and find the following: "In many adults, the benefits of aspirin exceed the risks (principally bleeding). For individuals aged ≥ 50 years without excess bleeding risk, we suggest low-dose daily aspirin (75 to 100 mg) (weak recommendation based on moderate confidence in effect estimates)."[1]

After reading the corresponding Users' Guide, you know that weak recommendations reflect the judgment of a specific group (eg, a guideline panel) that an individualized decision is necessary. You are curious, however, about the rationale for a weak recommendation for aspirin use and decide that this is a good opportunity to understand more about the GRADE (Grades of Recommendation Assessment, Development and Evaluation) approach to move from evidence to recommendations.

Direction and Strength of Recommendations

Like recommendations based on other evidence-rating systems, recommendations developed with the GRADE approach specify the direction of the recommendation (in favor or against the intervention) and are classified as strong or weak.

Strong recommendations apply to all or almost all patients and indicate that clinicians do not (if they are ready to trust the judgment of the panel) require thorough (or even cursory) review of the underlying evidence, or a discussion of benefits and risks with the patient.

Weak recommendations, in contrast, apply to the majority of patients, but not to everyone. To effectively use weak recommendations, clinicians need to understand and consider the key factors driving the direction and strength of the recommendation. Given that weak recommendations are usually sensitive to patients' values and preferences, they require that one be prepared to engage the patient in shared decision-making.

Developing GRADE Recommendations

GRADE provides a structure for assessing confidence in estimates from evidence summaries (i.e. systematic reviews) and, in the context of guidelines, also for moving from evidence to recommendations. Figure 28.1-1 presents the steps involved in developing a GRADE recommendation, including evidence synthesis and moving from evidence to recommendations. In the first step, guideline panelists formulate a clinical question, which involves specifying the target population, the intervention of interest, and an appropriate comparator. Conceptually, the final recommendation represents the answer to this question. Having formulated the question, guideline panelists select the relevant outcomes and rate their importance as critical, important, or not important for decision-making; the panel considers the critical or important outcomes in making their recommendation. Guideline panelists then consider all of the relevant studies, using existing systematic reviews or conducting their own reviews, including summaries that

provide estimates of both relative and absolute effect of the intervention vs. the comparator for these outcomes.

Using the evidence summaries, guideline panelists then rate the confidence in the effect estimates for each outcome as high, moderate, low, or very low. The GRADE working group has developed 2 standardized tabular formats for evidence summaries: the evidence profile and the summary-of-findings table.[2] The evidence profile provides a detailed record of the judgments made in the evaluation of the confidence in the estimates of effect and presents both relative and absolute estimates of the effect (Table 28.1-1). The summary-of-findings table provides a more concise summary of the same information (Table 28.1-2).

The next step is to move from the evidence to the recommendation. Panelists consider 4 factors as they ponder the direction and strength of a GRADE recommendation: (1) the overall confidence in effect estimates, (2) the balance between benefits and harms, (3) the uncertainty and variability in patients' values and preferences, and (4) resource considerations (Table 28.1-3). The recommendation and associated remarks are the final products of the GRADE process (Figure 28.1-1). As we will highlight, the process implies that although confidence in estimates of effect has an important influence on strength of recommendations, the 2 are quite separate: both high and low confidence in estimates can be associated with either strong or weak recommendations.

From Evidence to Recommendations

Overall Confidence in Effect Estimates

Confidence in estimates of effect (also known as quality of evidence) usually varies across outcomes. For instance, we are almost always more confident in estimates of the benefits of new interventions than in the frequency with which rare serious adverse effects occur. Therefore, systematic review authors and guideline panelists must present their confidence in estimates for each outcome. The GRADE guidelines also present an overall confidence in effect estimates, which represents the lowest rating among the outcomes considered critical to decision making.

High and moderate confidence in effect estimates reflects the panel's judgment that the effect estimates are sufficiently credible to support a particular recommendation. High and moderate confidence will justify strong recommendations in favor of or against a particular course of action if the desirable consequences (benefits) clearly outweigh the undesirable consequences (harms, burden, and costs). If desirable and undesirable consequences are closely balanced, however, even high confidence in estimates will not lead to a strong recommendation (Table 28.1-3).

For example, evidence from a systematic review of randomized trials with low risk of bias, consistent results across trials, and results applicable to the population (moderate confidence, due to imprecision) shows that, in people with nonvalvular atrial fibrillation, the use of warfarin, as compared with aspirin, reduces the risk of stroke but increases the risk of bleeding and the burden of treatment. In people at high risk of stroke (eg, congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or

transient ischemic attack [CHADS2] score of 3-6 points), the desirable consequences—a reduction of 40 strokes per 1000 patients treated for a year (95% CI, from 23 to 51 fewer)—clearly outweighs the undesirable consequences—an increase of 8 bleeds per 1000 patients treated for a year (95% CI, from 1 fewer to 10 more) and the need for laboratory monitoring and lifestyle changes. Consequently, a strong recommendation in favor of warfarin is appropriate.[3]

However, in persons at moderate risk of stroke (eg, CHADS2 score of 1 point), the balance between desirable and undesirable consequences is closer: the use of warfarin over aspirin reduces the risk of stroke by 9 events per 1000 patients treated for a year (95% CI, from 5 to 11 fewer) and produces the same increase in bleeding and burden of treatment as in people at high risk of stroke. In this situation, despite the moderate confidence in effect estimates, a weak recommendation in favor of warfarin is more appropriate.3

Low and very-low confidence in effect estimates reflect the guideline panel's judgment that there is considerable uncertainty regarding the outcomes associated with the alternative courses of action under consideration. In these circumstances, different attitudes toward uncertain benefits and harms are likely to lead to variability in what informed patients may choose, and guideline panels appropriately using the GRADE approach will most often issue weak recommendations.

For example, evidence from observational studies suggests that a diet rich in potassium might decrease cardiovascular risk. However, the confidence in effect estimates is low (coming from observational studies).[4] Although the intervention has no known adverse effects and almost no additional cost, some people are likely to judge that the uncertain benefit does not warrant the effort of changing their diet, while others will be prepared to do so. In this case, the uncertainty regarding the benefits of the intervention is responsible for variability in what informed patients will choose. Therefore, a weak recommendation (either for or against, depending on the panel's judgment about whether the majority of informed individuals, in response to the evidence, would change or not change their diet) is appropriate.[4]

In general, we should expect weak recommendations when the overall confidence in effect estimates is low or very low. Therefore, if you find a panel providing a strong recommendation on the basis of evidence warranting only low or very low confidence, you should be suspicious about that panel's judgment. Such situations warrant a careful look at why the recommendation was graded as strong.

Sometimes, however, guideline panels may appropriately offer strong recommendations despite low or very low confidence in effect estimates. Table 26-1 presents 5 paradigmatic situations in which strong recommendations may be warranted despite low confidence in key effect estimates.

For example, evidence warranting low confidence suggests that in males with androgen deficiency, testosterone supplementation improves well-being. However, indirect evidence (low confidence in effect estimates) suggests that testosterone might contribute to the spread of hormone-dependent cancers, including prostate cancer. If a guideline panel believes (as it reasonably might) that all or virtually all informed men with androgen deficiency and prostate cancer would opt not to take the risk of accelerated dissemination for the modest benefits of testosterone in well-being, a strong recommendation against testosterone will be appropriate.[5]

Balance Between Benefits and Harms

The balance between benefits and harms is a crucial determinant of both the direction and strength of a recommendation. If the net benefit (desirable vs. undesirable consequences) of an intervention vs. a comparator is small, guideline panels will generally issue weak recommendations (Table 28.1-3).

Two factors are involved in a panel's judgment of this balance: the magnitude of the benefits vs. harms (including the burden of treatment), and the relative importance that typical patients place in the benefits and harms.

For example, the use of warfarin in comparison with no therapy in people with nonvalvular atrial fibrillation and low risk of stroke (eg, CHADS2 score of 0) results in a reduction of 5 strokes per 1000 patients treated for a year (95% CI, 4 fewer to 6 fewer per 1000) but in an increase of 8 major bleeds per 1000 patients treated for a year (95% CI, from 1 more to 25 more per 1000)[3] and an associated burden of treatment. The decision

regarding net benefit is critically dependent on the relative value that patients place in the stroke, burden, and bleeding outcomes. Typical patients place a very high value in avoiding a stroke and its long-term consequences. The best available estimate suggests that for informed patients, avoiding a stroke is 3 times more important than avoiding a bleeding event.[6] If we take this factor into account, in patients with atrial fibrillation and low risk of stroke, the reduction of strokes is more important than the increase of bleeding: a reduction of 5 strokes per 1000 patients (multiplied by 3) vs. an increase of 8 major bleeds per 1000 patients (multiplied by 1). However, the net benefit is small and the burden of treatment substantial, particularly considering that patients with atrial fibrillation and low risk of stroke are typically young and without comorbidities. Consequently, most informed patients might choose to not use warfarin in this scenario; however, some might decide the opposite. A weak recommendation against warfarin in this circumstance is therefore appropriate.[3]

Clinicians using GRADE to evaluate specific recommendations should expect that guideline panels will not only present the absolute effects of interventions considered when balancing the benefits and harms of the intervention, but also explain the judgment of typical values and preferences they used to make the trade-off.

Uncertainty and Variability in Patients' Values and Preferences

Our discussion has emphasized the crucial role of values and preferences. If a guideline panel is very uncertain about patients' values and preferences, or believes these are extremely variable, a weak recommendation is likely (Table 28.1-3).

For example, vitamin K antagonists (VKAs) have been associated with fetal wastage, bleeding in the fetus, and teratogenicity if used after 8 weeks of gestation.[7] Women using long-term VKA treatment who are attempting pregnancy face the choice of performing frequent pregnancy tests and substituting parental anticoagulants for warfarin when pregnancy is achieved or replacing VKA with parental anticoagulants before conception is attempted. Both options have limitations. Small observational studies provide low confidence estimates suggesting that VKAs are safe during the first 6 to 8 weeks of gestation.[8] This provides some, but limited, reassurance to women who, considering the burden and cost of injections, would prefer to continue using warfarin until pregnancy is achieved. What informed patients prefer in this scenario remains uncertain, but it is likely to be variable. Consequently, a weak recommendation is appropriate, and even the direction is questionable. The 9th edition of the Antithrombotic Guidelines (AT9), issued a weak recommendation in favor of using VKA until pregnancy is achieved.[7]

Resource Considerations

The health budget—whether of a family, an organization, or a country—should be distributed fairly. Even if we focus only on interventions that lead to an important benefit on patient-important outcomes, depending on resource constraints, it may not be possible to offer them to all who might benefit. The existence of competing health needs and scarce resources suggests that to be optimally helpful, guideline panels may recommend against beneficial treatments when the gain is modest and the cost high.

Health economic analysis provides guidance for making these decisions. Economic analyses present the benefits and harms of the candidate interventions and their

associated resource utilization. The analyses facilitate selection of the interventions that offer the greater benefits in relation to the amount of resources used.

Strong recommendations reflect a panel's judgment that the benefits of the recommended course of action justify its resource use in all or almost all circumstances. If there is variability in the resource utilization across modes of delivery (eg, well-baby visits can be performed by physicians, nurses, and nurse-practitioners) or variability in resource availability (eg, high-income vs low-income countries) across the settings in which the recommendation will be applied, guideline panels are likely to issue weak recommendations (Table 28.1-3).

For example, the balance between the benefits and harms of using inhaled glucocorticoids in people with stable chronic obstructive pulmonary disease (COPD) is very close: evidence from randomized trials suggests that the use of inhaled glucocorticoids may reduce symptoms and the risk of exacerbations, but also may increase the risk of pneumonia.[9] This evidence has led some guideline developers to suggest the use of inhaled glucocorticoids in people with COPD, especially in those with more symptoms or with frequent exacerbations despite an optimal long-acting inhaled bronchodilator regimen.[10] However, high doses of the drug for long periods are required to achieve an appreciable effect, and the cost of the intervention is relatively high. A guideline panel from the World Health Organization considered that, for resource-limited settings, the small benefit of inhaled glucocorticoids does not justify its relative high cost in the majority of the circumstances, and consequently issued a strong recommendation against the intervention.11 If the target audiences of the guideline are in settings with different

availability of resources, a weak recommendation is more appropriate, or different recommendations are needed for the different settings.

Learning how resources issues impact recommendations will help you to assess whether and how the recommendation is relevant to your setting. Resource use issues are, however, not always included in guideline panel deliberations, and some panels may legitimately choose to not consider resources to develop their recommendations.

Clinical Scenario Resolution

You decide to explore the influence of the 4 factors discussed on the direction and strength of the recommendations. To do this, you look at the guideline text and Tables 28.1-1 and 28.1-2.

Overall Confidence in Effect Estimates

As the tables show, the guideline panelists considered 5 outcomes: mortality, myocardial infarction, stroke, major extracranial bleeding, and incidence of cancer. The first 4 were considered critical for decision making, while the last was considered important. The confidence in effect estimates was rated as high for myocardial infarction and major extracranial bleeding because the effect estimates came from a systematic review of randomized trials with low risk of bias, consistent results across trials, precise effects (narrow confidence intervals [CIs]), and results applicable to the population. For the outcomes of mortality and stroke, however, the 95% CI included both appreciable benefit and no benefit or harm, and hence panelists decreased the confidence to moderate. Finally, the confidence was judged as low for the outcome of incidence of cancer, since

the absolute estimates included a substantial benefit and a marginal effect (imprecision), and there was concern regarding the exclusion of 2 large studies from the original meta-analysis reporting this outcome (risk of bias). Typically, the overall confidence in effect estimates is the lowest rating among the outcomes considered as critical, which in this case is moderate because the outcome of incidence of cancer (rated as low) was considered important but not critical for decision making

Balance Between Benefits and Harms

Considering the baseline risk of individuals at 60 years of age and at average risk for coronary artery disease (10% to 20% over 10 years) and at average risk of malignancy (approximately 5%), the use of aspirin prevents 6 deaths per 1000 (95% CI, from 12 to 0 fewer), 19 nonfatal myocardial infarctions (95% CI, from 26 to 12 fewer) and 6 new cancers (95% CI, from 10 to 1 fewer) over a 10-year period. However, it produces 16 extracranial major bleeds (95% CI, from 7 to 20 more) over the same period (Tables 28.1-1 and 28.1-2). It is likely that most informed patients will place a higher value on avoiding mortality, vascular events, and cancer than on the possibility of bleeding, and the use of aspirin will be perceived as a net benefit by informed patients. The absolute magnitude of this benefit, however, is small, since all of the events considered are infrequent in primary prevention populations.

Uncertainty and Variability in Patients' Values and Preferences

The net benefit of aspirin is very small in absolute terms. Some informed patients will be willing to tolerate the long-term medication use to gain a small reduction in the risks of

death, vascular events, and cancer and will tolerate the risk of bleeding. Others, however, are likely to consider that the effect is not of sufficient magnitude to warrant the inconvenience and small risk of bleeding, and consequently, they will not be willing to use aspirin. Therefore, we can expect variability in what informed patients may choose.

Resource Considerations

Resource utilization was not explicitly considered in this recommendation. However, with aspirin's minimal expense, cost issues are unlikely to play an important part in this recommendation.

Integrating the Factors

The benefits of aspirin (reduction of mortality, myocardial infarction, and incidence of cancer) outweigh the risks (increase of bleeding and burden), and hence, the guideline panel issued a recommendation in favor of aspirin.

Regarding the strength of the recommendation, the moderate overall confidence in effect estimates may have warranted a strong recommendation. However, the small magnitude of the benefit and the likely variability in patients' values and preferences appropriately led the panel to grade the recommendation as weak.

Now that you understand the panel's rationale for the recommendation, you are in a position to engage in shared decision making with the 60-year-old man considering use of aspirin for primary prevention.

Table and Figures

Figure 28.1-1. Steps Involved in Developing a GRADE Recommendation

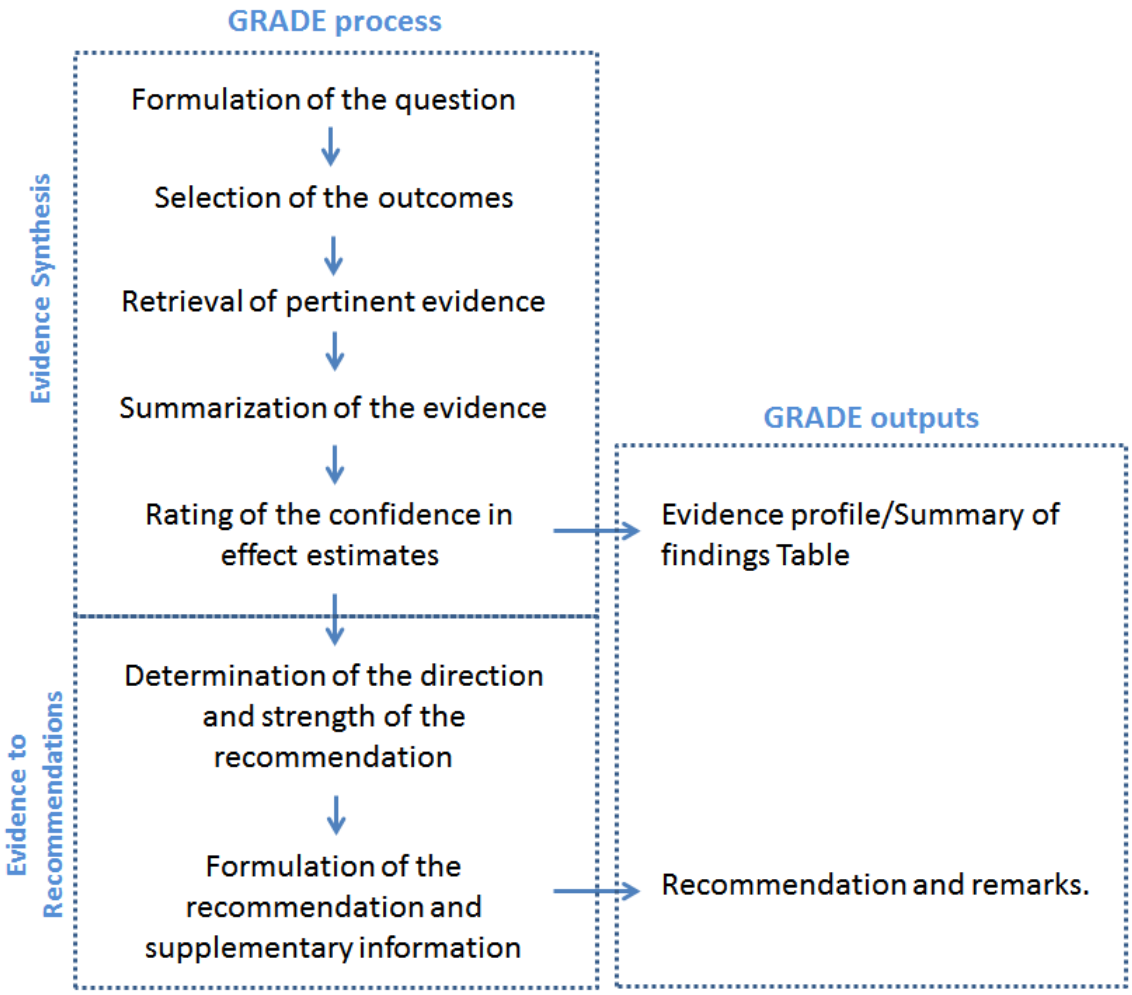


Table 28.1-1. Evidence Profile: Aspirin (75 to 100 mg) Compared With No Aspirin in the Primary Prevention of Cardiovascular Disease

Quality Assessment							Effect		Confidence	Importance
No. of Studies (participants)	Study Design	Risk of Bias	Inconsistency	Indirectness	Imprecision	Other Considerations	RR (95% CI)	Absolute Risk Over 10 Years		
Mortality										
9 (100,076)	RCTs	Not serious	Not serious	Not serious	Serious[b]	Not serious	RR, 0.94 (0.88-1.00)	6 fewer deaths per 1000 (from 12 fewer to 0 fewer)	⊕⊕⊕O MODERATE	CRITICAL
Myocardial infarction (MI). Nonfatal events										
9 (100,076)	RCTs	Not serious	Not serious	Not serious	Not serious	Not serious	RR 0.80 (0.67-0.96)	17 fewer MI per 1000 (from 27 fewer to 3 fewer)	⊕⊕⊕⊕ HIGH	CRITICAL
Stroke. Includes nonfatal ischemic and hemorrhagic strokes										
9 (95,000)	RCTs	Not serious	Not serious	Not serious	Serious[b]	Not serious	RR, 0.95 (0.85-1.06)	No significant difference 3 fewer strokes per 1000 (from 10 fewer to 4 more)	⊕⊕⊕O MODERATE	CRITICAL
Major extracranial bleeding										
6 (95,000)	RCTs	Not serious	Not serious	Not serious	Not serious	Not serious	RR, 1.54 (1.30-1.82)	16 more bleeds per 1000 (from 7 more to 20 more)	⊕⊕⊕⊕ HIGH	CRITICAL
Cancer incidence										
6 (35,535)	RCTs	Serious[c]	Not serious	Not serious	Serious[d]	Not serious	HR, 0.88 (0.80-0.98)	6 fewer cancers per 1000 (from 10 fewer to 1 fewer)	⊕⊕OO LOW	IMPORTANT

Abbreviations: HR, hazard ratio; MI, myocardial infarction; RCTs, randomized clinical trials; RR, relative risk.
[a]Baseline risks included in the table are from medium-risk patients according to the Framingham Score. These risks correspond to the patient presented in the opening scenario.
[b]The 95% CI included both appreciable benefit and no benefit or harm.
[c]Two large studies were arbitrarily excluded from the original meta-analysis.
[d]The absolute estimates included a substantial benefit and a marginal effect.

Adapted from Spencer et al.[1]

Table 28.1-2. Summary of Findings Table: Aspirin (75 to 100 mg) Compared With No Aspirin in the Primary Prevention of Cardiovascular Disease

Outcome	No. of Patients (Studies)	Relative Effect (95% CI)	Anticipated Absolute Effects Over 10 Years		Confidence in Effect Estimates (GRADE)
			Without Aspirin	Risk Difference With Aspirin (95% CI)	
Mortality	100 076 (9)	RR, 0.94 (0.88-1.00)	100 deaths per 1000	6 fewer deaths per 1000 (from 12 fewer to 0 fewer)	⊕⊕⊕O MODERATE due to imprecision[b]
Myocardial infarction (MI) (nonfatal events)	100 076 (9)	RR 0.80 (0.67-0.96)	83 MI per 1000	17 fewer MI per 1000 (from 27 fewer to 3 fewer)	⊕⊕⊕⊕ HIGH
Stroke (nonfatal ischemic and hemorrhagic strokes)	95 000 (9)	RR, 0.95 (0.85-1.06)	65 strokes per 1000	No significant difference 3 fewer strokes per 1000 (from 10 fewer to 4 more)	⊕⊕⊕O MODERATE due to imprecision
Major extracranial bleeding	95 000 (6)	RR, 1.54 (1.30-1.82)	24 bleeds per 1000	16 more bleeds per 1000 (from 7 more to 20 more)	⊕⊕⊕⊕ HIGH
Cancer (incidence)	35 535 (6)	HR, 0.88 (0.80-0.98)	50 cancers per 1000	6 fewer cancers per 1000 (from 10 fewer to 1 fewer)	⊕⊕OO LOW due to imprecision and risk of bias[c,d]

Abbreviations: HR, hazard ratio; GRADE, Grades of Recommendation Assessment, Development and Evaluation; RR, relative risk.
[a]Baseline risks included in the table are from medium-risk patients according to the Framingham Score. These risks correspond to the patient presented in the opening scenario.
[b]The 95% CI included both appreciable benefit and no benefit or harm.
[c]Two large studies were arbitrarily excluded from the original meta-analysis.
[d]The absolute estimates included a substantial benefit and a marginal effect.

Adapted from Spencer et al.[1]

Table 28.1-3. Factors That Determine the Strength of the Recommendation

	A Strong Recommendation May Be Justified If	In General, We Should Expect a Weak Recommendation When
Overall confidence in effect estimates	There is high or moderate confidence in effect estimates (or in special circumstance when the confidence is low or very low) **AND**	There is low or very-low confidence in effect estimates **OR**
Balance between benefits and harms	The benefits clearly outweigh the harms or vice versa **AND**	The balance between benefits and harms is close **OR**
Uncertainty and variability in patients' values and preferences	All or almost all fully informed patients will make the same choice **AND**	There is variability or uncertainty in what fully informed patients may choose **OR**
Resource considerations (optional)	The cost of the intervention is clearly justified (or not) in all or almost all the circumstances	The cost of the intervention may not be justified in some circumstances

References

1. Spencer FA, Guyatt GH. Aspirin in the primary prevention of cardiovascular disease and cancer. In: UpToDate, Basow, DS (Ed), UpToDate, Waltham, MA, 2013. (http://www.uptodate.com/contents/aspirin-in-the-primary-prevention-of-cardiovascular-disease-and-cancer. Date of access: April 10th, 2014)
2. Guyatt G, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines 1. Introduction GRADE evidence profiles and summary of findings tables. 2011 Apr;64(4):383-94.
3. You JJ, Singer DE, Howard PA, Lane DA, Eckman MH, Fang MC, Hylek EM, Schulman S, Go AS, Hughes M, Spencer FA, Manning WJ, Halperin JL, Lip GY. Antithrombotic therapy for atrial fibrillation: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines.Chest. 2012 Feb;141(2 Suppl):e531S-75S.
4. WHO Guideline: Potassium intake for adults and children. Geneva, World Health Organization (WHO), 2012. (http://www.who.int/nutrition/publications/guidelines/potassium_intake_print version.pdf. Date of access: April 10th, 2014)
5. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force, Endocrine Society. Testosterone therapy in

men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010 Jun;95(6):2536-59.

6. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, Nagpal S, Cox JL.Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ. 2001 Nov 24;323(7323):1218-22.

7. Bates SM, Greer IA, Middeldorp S, Veenstra DL, Prabulos AM, Vandvik PO. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012 Feb;141(2 Suppl):e691S-736S.

8. Schaefer C, Hannemann D, Meister R, et al. Vitamin K antagonists and pregnancy outcome. A multi-centre prospective study. Thromb Haemost. 2006;95(6):949-957.

9. Yang IA, Clarke MS, Sim EH, Fong KM. Inhaled corticosteroids for stable chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2012 Jul 11;7:CD002991.

10. Erbland ML. Role of inhaled glucocorticoid therapy in stable COPD. In: UpToDate, Basow, DS (Ed), UpToDate, Waltham, MA, 2013. (http://www.uptodate.com/contents/role-of-inhaled-glucocorticoid-therapy-in-stable-copd#H14. Date of access: April 10th, 2014)

11. WHO Guideline: Prevention and control of NCDs: Guidelines for primary health care in low-resource settings. Geneva, World Health Organization

(WHO), 2012. (http://apps.who.int/iris/bitstream/10665/76173/1/9789241548397_eng.pdf. Date of access: April 10th, 2014)

12. Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, Yazdanpanah Y, Beigel J, Chotpitayasunondh T, Del Mar C, Farrar J, Tran TH, Ozbay B, Sugaya N, Fukuda K, Shindo N, Stockman L, Vist GE, Croisier A, Nagjdaliyev A, Roth C, Thomson G, Zucker H, Oxman AD; WHO Rapid Advice Guideline Panel on Avian Influenza. WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. Lancet Infect Dis. 2007 Jan;7(1):21-31
13. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, Lynch DA, Ryu JH, Swigris JJ, Wells AU, Ancochea J, Bouros D, Carvalho C, Costabel U, Ebina M, Hansell DM, Johkoh T, Kim DS, King TE Jr, Kondoh Y, Myers J, Müller NL, Nicholson AG, Richeldi L, Selman M, Dudden RF, Griss BS, Protzko SL, Schünemann HJ; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011 Mar 15;183(6):788-824
14. Freedman AS, Lister A, Connor RF. Management of gastrointestinal lymphomas. In: UpToDate, Basow, DS (Ed), UpToDate, Waltham, MA, 2013. (http://www.uptodate.com/contents/management-of-gastrointestinal-lymphomas. Date of access: April 10th, 2014)

15. Bates SM, Greer IA, Middeldorp S, Veenstra DL, Prabulos AM, Vandvik PO. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012 Feb;141(2 Suppl):e691S-736S.

16. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM; Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2010 Jun;95(6):2536-59.

Authors' contributions:

Dr. Neumann contributed to the conception of the chapter, data collection and drafting of the article. All the coauthors contributed to the conception of the chapter and drafting of the article

Chapter 4: Experiences with a novel policy for managing conflicts of interest of guideline developers: A descriptive qualitative study

Ignacio Neumann, Renee Karl, Aman Rajpal, Elie A. Akl and Gordon H. Guyatt.

*This paper was published in a peer-reviewed journal. Reference: Neumann I, Karl R, Rajpal A, Akl EA, Guyatt GH. Experiences with a novel policy for managing conflicts of interest of guideline developers: A descriptive qualitative study. Chest. 2013 Aug;144(2):398-404. doi: 10.1378/chest.12-2390. (http://journal.publications.chestnet.org/article.aspx?articleid=1573740)

Abstract

Background: The executive committee of the Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (AT9) developed a novel policy for managing conflicts of interest (COIs): Methodologists bore primary responsibility for each chapter, there was equal emphasis on intellectual and financial COI, and content experts with COIs participated, but with restrictions for recommendations on which they had conflicts. The objective of this study was to explore the experiences of the methodologists and content experts with the COI policy after its implementation.

Methods: One investigator conducted two rounds of semi-structured interviews with the methodologist and the leading content expert of each chapter until data saturation was achieved. Two investigators analyzed the transcripts of the interviews in duplicate using an immersion-crystallization approach. We also conducted member checking.

Results: We interviewed 15 participants and presented the results to the remaining four for verification. In comparison with their views expressed prior to AT9 development, methodologists remained more positive about the policy than content experts. Six of 10 content experts expressed a more positive view than prior to participation in the AT9 process. The other four content experts remained skeptical, especially regarding the emphasis on intellectual COI. The restrictions of the policy on conflicted individuals were not fully implemented.

Conclusions: After its implementation, some content experts were more favorable to the policy, but some retained major reservations. The influence of the policy on

recommendations may have been more through the leading role of the methodologists than exclusion of conflicted participants in making recommendations.

Background

Clinical practice guidelines can improve health-care efficiency, promote medical interventions of proven benefit, and discourage harmful care.[1-3] Guideline panelists often report conflict of interest (COI), in particular, ties with the pharmaceutical industry.[4-13] Nonfinancial conflicts, including, for instance, professional and institutional loyalties and, in particular, intellectual COI related to prior writings and other work, may also influence recommendations.[14,15] Simply disclosing potential COIs may not minimize their impact, and excluding conflicted content experts altogether may negatively affect the quality of guidelines. New strategies that limit the impact of COI while allowing the participation of content experts are needed to ensure the optimal quality of clinical guidelines.

The executive committee of the Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (AT9) developed a novel policy designed to use experts' input while limiting the impact of their COIs on recommendations.[16] Initially, the Health and Sciences Policy Committee of the American College of Chest Physicians (ACCP) rejected 13 of 150 nominees on the basis of excessive financial COI. Among the accepted panelists, methodologists free of COIs ("methodologists" in this article) bore primary responsibility for the recommendations in the 14 recommendation chapters. A content area expert with, potentially, both financial and intellectual COIs ("content experts" in this article) had

input into preparing, summarizing, and interpreting the evidence. Like other panelists, however, the intent was to exclude them from the deliberations that ultimately determined the direction and strength of recommendations on which they had conflicts (Table 1).

AT9 used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to guidelines, which includes rating confidence in estimates of effect and grading strength of recommendations as strong or weak.[17] The COI policy may have influenced the direction of the recommendations (for or against an intervention), the strength of the recommendations (strong, indicating appropriate for all or almost all patients, or weak, appropriate for the majority but not all), or the ratings of the confidence in estimates of effect (high, moderate, or low in the AT9/GRADE system). The hypothesis when instituting the process was that more stringent application of GRADE would sometimes lead to lower confidence in estimates, and that this, along with reduced COIs, would lead to fewer strong recommendations.

Prior to the implementation of the policy, a qualitative study found that the methodologists and content experts were uneasy regarding their respective roles. Although methodologists believed that the COI policy would ensure more rigorous guidelines, some content experts expressed concerns that methodologists' lack of content expertise might compromise the quality of the guidelines.[18]

The issue of intellectual COI plays a major role in the ensuing results and discussion. The authors would like to acknowledge their own intellectual conflict: Three of the five of us (Drs Neumann, Akl, and Guyatt) are clinical epidemiologists with a strong investment in the GRADE process for developing clinical practice guidelines. Ms Karl and Dr Rajpal declare no conflicts of interest.

Objectives

This study has three main objectives:

1. Compare the experiences of the AT9 methodologists and of the content experts regarding the implementation of the new COI policy.

2. Compare those experiences with their initial reactions to the policy.

3. Explore how the policy impacted the process of developing the guideline and perceptions of its impact on the final product.

Materials and Methods

Setting and Study Population

We invited participation from nine of the 10 methodologists (one was excluded due to his involvement in this study) and the 10 leading content experts who were involved in the development of AT9 and participated in the previous qualitative study.[18] All participants gave their informed consent at the beginning of the interviews. The Institutional Review Board of the McMaster University approved the study (11-129-S).

Data Collection

One investigator with no other role in AT9 conducted two rounds of semi-structured interviews with participants in person, when possible, or by teleconference (the first round during the guidelines development process and the second 5 months after the guidelines final submission). The interviews were guided by a set of predefined questions based on the findings of the qualitative study conducted before the implementation of the policy[18] and continued until participants had addressed all the previously identified

questions and were satisfied that they had expressed their relevant views. The interviews were audio recorded and one investigator transcribed them to electronic text files. The interviewer reviewed all transcripts to ensure their accuracy.

Data Analysis

Two investigators analyzed the transcripts independently using an immersion-crystallization approach.[19] This approach is an iterative, systematic process of reviewing transcripts, identifying segments representing themes, assigning codes to the specific themes, and organizing the themes into categories. The process is repeated until no new themes emerge. The two investigators identified consistent themes through their independent analysis.

Member Checking

To verify our findings, we invited all participants to review and comment on a draft report of the results. We also sent this synthesis to the panelists who were not able to participate in the interviews and requested comments and, in particular, any disagreement with the synthesis.

Results

Description of Context and Participants

We interviewed eight of the nine methodologists and seven of the 10 content experts (Tables 2, 3). Other concurrent commitments prevented the participation of one

methodologist and three content experts. These four individuals subsequently provided feedback and comments on the summary of the results of the interviews.

The methodologists, eight of whom were medical doctors with clinical experience and one a biostatistician, were typically less advanced in their academic career than content experts. Five content experts had financial and intellectual COIs, one had financial COIs only, and one had intellectual COIs only; the three without COIs participated in chapters not related to their research focus (Table 2).

Themes

We first present two themes about the overall impact of the policy on the guideline process and on the final recommendations. We then present the findings related to two critical aspects of the novel policy: restrictions in participation of conflicted individuals and emphasis in intellectual COIs. Finally, we explore the relationship between methodologist and content experts.

Impact of the Policy on the Guidelines Process:

Methodologists were much more positive about the policy than content experts. One stated: "I think that it is a great policy. I think that there is a big need for a policy like this." Methodologists believed that the consideration of intellectual COIs was an important innovation and that the implementation of concrete rules increased transparency. However, three methodologists expressed concern that the guidelines may be less rich because of content expert disengagement and the exclusion of some of the

field's leaders. Methodologists believed that their initial worries that less seniority might impair their function proved groundless.

Among content experts, there were disagreements regarding the value of the policy. Six content experts (five with both intellectual and financial COIs, one with financial COI only) expressed a favorable opinion of the policy, whereas the other four (three with both intellectual and financial COIs, one with intellectual COI only) were skeptical. The skeptical believed that the process of identifying COIs was overly cumbersome and that decisions about who was excluded were somewhat arbitrary. In addition, some believed that in some chapters the restrictions on conflicted individuals gave methodologists and other unconflicted panelists too much power. As one content expert said: "In some chapters few panel members were able to vote, so methodologists and front line clinicians had a lot of power. Not sure if this was the best way….some experts who I am sure had a lot to contribute felt so cut off by the policy that they never really did become fully engaged." Despite these negative views, all content experts valued the addition of an unconflicted methodologist (even those who believed the methodologist had too much power) and considered this an important innovation.

Impact of the Policy on the Guidelines Product:

Perceptions of the impact that the policy had on the final recommendations varied by chapter. In some chapters, participants perceived that the policy had little impact on the guidelines. In others, methodologists and content experts suggested that the implementation of the policy resulted in recommendations moving from strong to weak (in AT9, the proportion of recommendations that were strong was 33% compared with

64% in the previous iteration). One methodologist pointed out: "There are often situations where the quality is not that great, but people still want strong recommendation, but I think the COI policy prevented that from happening to a certain degree…fewer strong recommendations based on low quality evidence."

Content experts believed that a trade-off remained between managing COIs, adhering to GRADE methodology, and allowing maximum participation of knowledgeable but conflicted individuals. They also expressed concerns about how clinicians will respond to recommendations formerly being strong now being weak and quality of evidence previously classified as high or moderate now being classified as moderate or low. Content experts believed that clinicians' response to the guidelines will be the ultimate judge of the effectiveness of the new process: "I think it will be very interesting to see how the medical community accepts them. There will be many changes compared to prior versions so I think there may be some discomfort about how a panel can review the same evidence and come to different conclusions. So I don't know how they will be accepted and of course there will be a competition with other guidelines that other organizations have to offer. So I am happy with them but I don't know how they will be received."

Restrictions on Participation of Conflicted Individuals:

Six of the eight methodologists and six of the seven content experts interviewed perceived that the restriction on participation in the discussions of conflicted individuals was followed loosely. One methodologist stated: "I have been a little bit lenient, I would say because we kept hammering the recommendations more or less until everybody was

happy…So we kept allowing everybody to at least keep commenting on all the topics." Typically, conflicted panelists were allowed to participate to some degree in the discussions of the quality of the evidence and the direction and strength of the recommendations.

All the participants indicated that the rule banning conflicted individuals from voting had limited impact on the guidelines, since formal voting occurred in only a small proportion of recommendations. One content expert noted that the self-awareness of the COI may have led conflicted individuals to restrain themselves from strong advocacy.

The Emphasis on Intellectual COI:

The introduction of intellectual COI was the most controversial element of the new policy. All the methodologists believed that intellectual COI was a potential source of bias. As one methodologist commented, "Sometimes people who have intellectual COI may cause more bias than financial COI, because when you do the research you obviously believe that what you did is right."

In contrast, only five of 10 content experts believed that intellectual COI could pose a source of bias. One content expert commented: "I think intellectual COI is just crap. It's artificial and it's just for show." Another stated, "I think a lot of people have difficulty with the assumption that just because I write about something I am not able to make rational decisions…So when you set up things like intellectual COI it makes your intellectual contributions seems like a bad thing, like an evil influence."

Even content experts who believed that intellectual COI was a legitimate consideration believed that the policy was applied inconsistently and seemed somewhat arbitrary. They

also believed that the policy underestimated methodologists' intellectual COI, since it did not take into account their professional investment in GRADE methodology and authorship of systematic reviews.

Relationship Between Methodologist and Content Experts:

The relationship between the methodologist and the content expert was vital to the success of the process. In 12 of the 14 chapters, even if there were initial problems, methodologists and content experts established mutually respectful and productive relationships. There were two instances of conflicts. In these cases, content experts stated that methodologists applied GRADE methodology too rigidly and did not give sufficient attention and/or value to clinical experience. Methodologists believed that the content experts had strong opinions not supported by evidence. In some cases, content experts' frustration with their perceived "demotion" was not a source of immediate conflict but impacted the tone of the relationship. In the two chapters in which there were conflicts between methodologists and content experts, those involved voiced an unwillingness to participate in subsequent ACCP antithrombotic guidelines.

Discussion

Before participation, content experts were uniformly skeptical regarding the new policy.[18] After the process of guideline development was complete, six of 10 content experts were much more positive about the new structure, and the other four remained negative. Methodologists in general were more positive than content experts regarding the new policy and its consequences before and after the experience. Participants believed that the

ultimate impact of the policy on the recommendations varied across chapters. Participants with previous experience in the ACCP antithrombotic guidelines generally believed that the new AT9 process increased the number of weak recommendations.

Consistent with the discussion of the controversy in the literature, the introduction of intellectual COI as a source of bias was the most controversial element of the new policy.[20-22] In the study conducted before the implementation, all methodologists and six of 10 content experts agreed that intellectual COI might be a source of bias, whereas four of 10 content experts were skeptical. The experience of participating in AT9 did not substantially change these opinions. Indeed, content experts made stronger statements against the concept of intellectual COI in the interviews after the process. Additionally, some content experts were concerned that the policy that focused on COIs related to publication of original data may have underemphasized potential intellectual COIs of methodologists who had written relevant systematic reviews.

There is an apparent contradiction in some findings. Content experts perceived that restrictions were seldom applied and that experts with COIs were typically allowed to participate in the development of recommendations. But at the same time, they believed that restrictions produced feelings of exclusion and the disengagement of some panelists. The apparent contradiction may be related to loss aversion[23,24]: the potential for exclusion or the relatively few instances when the restrictions were actually applied (ie, content experts were excluded) may have had a disproportionate impact on feelings and perceptions.

The strengths of this study include strategies to ensure the trustworthiness of our findings. The data collection and analysis were conducted by investigators unrelated to the

development of AT9, the analysis was done independently and in duplicate, after the initial analysis we conducted follow-up interviews with all the participants, and we conducted member checking.

Our study has limitations. We restricted our interviews to methodologists and leading content experts. Since most of the other panelists were content experts, their views may have been similar to those expressed by the leading content experts.

The results of this study raise important issues in the management of COIs in clinical practice guidelines. The implications of one finding seem clear: All participants, even content experts who chafed under a perceived demotion, agreed that having methodologists in a leadership role improved the quality of the guidelines. Particularly with the increasing popularity and endorsement of the rigorous GRADE process for developing guidelines,[17] a leadership role for methodologists in clinical practice guidelines appears advisable.

What inference to make from the continued dissatisfaction with the process among most content experts is less clear. On the one hand, one might conclude that this represents a failure of the process, and that other strategies for managing COI—such as ensuring representation of diverse viewpoints[20]—are likely to prove superior. However, even conflicted individuals appear to have had substantial input into the recommendations. Loss of power and perceived prestige might inevitably lead to dissatisfaction and sometimes hostility, even if the new process resulted in higher-quality and more scientifically robust recommendations.

Attitudes toward intellectual COIs could potentially have a major influence on which interpretation one takes. The notion of intellectual COI elicits antagonism among many

commentators, including some of the content experts who participated in this study. Whether this reaction is restricted to content area experts who would be the subjects of the application of intellectual COI policies remains uncertain. Although anecdotal evidence of intellectual COI abounds, systematic study is very limited.

The experience with the COI policy implemented in AT9 provides at least two lessons. First, all agreed that unconflicted methodologists as chapter editors provided a positive innovation and perhaps the most effective safeguard against undue influence of COI on recommendations. Second, restrictions to conflicted panelists were difficult to fully implement and, despite the limited implementation, resulted in dissatisfaction among content experts. The final standard for the success of the new policy is whether it did or did not produce higher-quality guidelines manifested in recommendations more consistent with the available evidence and the values and preferences of the individuals to whom the recommendations will be applied. It is not altogether clear, however, how one should make this judgment, and who should make it. The results of this study suggest that the methodologists involved would deliver a resounding “yes” to this question; at least some of the content experts would answer in the negative. These considerations highlight the urgent need for continued systematic study to optimize the process of developing clinical practice guidelines, with the quality of the final product (ie, the guideline recommendations) being the major outcome of interest.

Tables

Table 1 —Key Features of the New Policy for Managing COIs Implemented in AT9

Key elements of the policy implemented in AT9:

1. Panelists self-reported their financial and intellectual COI using a standardized Excel sheet along with instructions and definitions for the types (financial and intellectual) of COI. Although there were instances in which the executive became aware of unreported conflicts and raised these with the panelists, no formal or systematic checks of self-reports were undertaken

2. A methodologist free of financial or intellectual conflicts bore primary responsibility for each of the AT9 chapters.

3. There was equal emphasis on intellectual and financial COI.

4. Experts with too many COI were excluded from the development of AT9.

5. Rules, distributed and explained to participants in advance, stipulated that only panel members without primary conflicts could be involved in the development of the recommendation. Content experts with a conflict of interest could contribute to preparing, summarizing, and interpreting the evidence but not to vote or participate on the discussions that ultimately determine the direction and strength of recommendations on which they are conflicted. Methodologists bore the responsibility of enforcing the compliance with the restriction imposed to conflicted panelists

Definitions used in AT9:

Financial COI: Include consultancies, advisory board membership, and the like from industry.

Intellectual COI: Academic activities that create the potential for an attachment to a specific point of view that could unduly affect an individual's judgment about a specific recommendation. Such activities include receipt of a grant or participation in research or commentary directly related to that recommendation.

Table 2 —Description of AT9 Panels, Methodologists, and Content Experts

The AT9 Guidelines included fourteen chapters with graded recommendations. The panel of each chapter consisted of a methodologist, a number of content experts and a frontline clinician. The methodologist had expertise in health research methods, had no conflicts of interest and led the chapter as its editor ("methodologist" in this paper). One of the content experts, who may have had financial and intellectual COI for some recommendations, served as the "deputy editor" ("content expert" in this paper)

Methodologists

Eight methodologists were medical doctors with clinical experience and one was a biostatistician. All had formal training in health research methodology. Six worked in North America and three in Europe. They tended to be less advanced in their academic career than content experts and for eight of nine this was the first experience working in the ACCP antithrombotic guidelines. None had any COI with AT9 recommendations

Leading Content Experts

All the content experts were medical doctors with formal training in a subspecialty relevant to their respective chapter(s). Eight worked in North America, one in United Kingdom and one in Australia. They were generally more advanced in academic career than methodologists. Six of ten had participated in previous iterations of the ACCP antithrombotic guidelines. Seven of ten declared a relevant COI (intellectual or financial) with at least one recommendation of their respective chapter. The three content experts without COI participated in chapters not related to their research focus.

Table 3 —Characteristics of the Included Methodologists and Content Experts

	Methodologists (n=9)	**Content Experts (n=10)**
Gender	*2 Females, 7 Males*	*2 Females, 8 Males*
Training	*8 Medical Doctors, 1 Biostatistician*	*10 Medical Doctors*
Postgraduate degrees	*PhD level: 3* *MSc level: 6*	*PhD level: 2* *MSc level: 3*
Country of residency	*USA: 4* *Canada: 2* *Spain: 1* *Norway: 1* *Switzerland: 1*	*USA: 3* *Canada: 5* *UK: 1* *Australia: 1*
Academic career	*Professor: 1* *Associate Professor: 7* *Assistant Professor or lower: 1*	*Professor: 7* *Associate Professor: 2* *Assistant Professor or lower: 1*
Participation in previous iteration (AT8)	*Yes: 1* *No: 8*	*Yes: 6* *No: 4*
Conflict of interest with at least one recommendation	*Yes: 0* *No: 9*	*Yes: 7* *No: 3*

References

1. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999 Feb 20;318(7182):527-30.
2. Bahtsevani C, Udén G, Willman A. Outcomes of evidence-based clinical practice guidelines: a systematic review. Int J Technol Assess Health Care. 2004 Fall;20(4):427-33.
3. Lohr KN, Eleazer K, Mauskopf J. Health policy issues and applications for evidence-based medicine and clinical practice guidelines. Health Policy. 1998 Oct;46(1):1-19.
4. Detsky AS. Sources of bias for authors of clinical practice guidelines. CMAJ. 2006 Oct 24;175(9):1033, 1035.
5. Johnson K. Medical societies scramble to include disclosure on clinical guidelines. CMAJ. 2011 May 17;183(8):E443-4.
6. Kahn JM, Rubenfeld GD. Financial disclosures in clinical practice guidelines. Crit Care Med. 2010 Aug;38(8):1755-6
7. Holloway RG, Mooney CJ, Getchius TS, Edlund WS, Miyasaki JO. Invited Article: Conflicts of interest for authors of American Academy of Neurology clinical practice guidelines. Neurology. 2008 Jul 1;71(1):57-63.
8. Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. PLoS One. 2011;6(10):e25153. Epub 2011 Oct 19.

9. Cosgrove L, Bursztajn HJ, Krimsky S, Anaya M, Walker J. Conflicts of interest and disclosure in the American Psychiatric Association's Clinical Practice Guidelines. Psychother Psychosom. 2009;78(4):228-32.

10. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011 Oct 11;343:d5621. doi: 10.1136/bmj.d5621.

11. Mendelson TB, Meltzer M, Campbell EG, Caplan AL, Kirkpatrick JN. Conflicts of interest in cardiovascular clinical practice guidelines. Arch Intern Med. 2011 Mar 28;171(6):577-84.

12. Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA. 2002 Feb 6;287(5):612-7.

13. Graham R, Mancher M, Dianne MW, Green_eld S, Steinberg E. Clinical Practice Guidelines We Can Trust. 1st ed. Institute of Medicine of the National Academies; 2011.

14. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

15. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012 Jul;65(7):725-33.

16. Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, Lewis SZ, Nathanson I, Jaeschke R, Schünemann H. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010 Jun 1;152(11):738-41

17. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ.GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011 Apr;64(4):383-94. Epub 2010 Dec 31.

18. Akl EA, Karl R, Guyatt GH. Methodologists and context experts disagreed regarding managing conflicts of interest of clinical practice guidelines panels. J Clin Epidemiol. 2012 Jul;65(7):734-9.

19. Strauss AL. Qualitative analysis for social scientists. New York: Cambridge University Press; 1987.

20. Sniderman AD, Furberg CD. Pluralism of viewpoints as the antidote to intellectual conflict of interest in guidelines. J Clin Epidemiol. 2012 Jul;65(7):705-7

21. Bion J. Financial and intellectual conflicts of interest: confusion and clarity. Curr Opin Crit Care. 2009 Dec;15(6):583-90.

22. Poole D. When does a point of view become an intellectual conflict of interest? Crit Care Med. 2008 May;36(5):1688

23. Kahneman D. A perspective on judgment and choice: mapping bounded rationality. Am Psychol. 2003 Sep;58(9):697-720

24. McGraw AP, Larsen JT, Kahneman D, Schkade D. Comparing gains and losses. Psychol Sci. 2010 Oct;21(10):1438-45.

Authors' contributions:

Dr Neumann had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Dr Neumann: contributed to the conception and design of the study, data collection, analysis and interpretation of the data, and drafting of the article.

Ms Karl: contributed to the conception and design of the study, analysis and interpretation of the data, and drafting of the article.

Dr Rajpal: contributed to transcribing the interviews and writing and revising the article.

Dr Akl: contributed to the conception and design of the study, analysis and interpretation of the data, and drafting of the article.

Dr Guyatt: contributed to the conception and design of the study, analysis and interpretation of the data, and drafting of the article.

Chapter 5: Low anonymous voting compliance with the novel policy for managing conflicts of interest implemented in the ninth version of the American College of Chest Physicians antithrombotic guidelines

Ignacio Neumann, Elie A. Akl, Macarena Valdes, Stephanie Bravo, Silvana Araos, Victor Kairouz, Holger Schünemann, and Gordon H. Guyatt,

*This paper was published in a peer-reviewed journal. Reference: Neumann I, Akl EA, Valdes M, Bravo S, Araos S, Kairouz V, Schünemann H, Guyatt GH. Low anonymous voting compliance with the novel policy for managing conflicts of interest implemented in the 9th version of the American College of Chest Physicians antithrombotic guidelines. Chest. 2013 Oct;144(4):1111-6. doi: 10.1378/chest.13-0169.

(http://journal.publications.chestnet.org/article.aspx?articleid=1690823)

Abstract

Background: The executive committee of the Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (AT9) developed a strategy to limit the impact of conflict of interest (COI) on recommendations. This policy excluded conflicted panelists from voting on recommendations with which they had conflicts. The objective of the study is to explore the compliance of the attendees of the AT9 final conference.

Methods: We conducted a survey and reviewed public declarations of COI of all the final AT9 conference attendees. For each of the controversies on which voting occurred (nine of 628 total recommendations), we estimated the compliance with COI policy as the proportion of attendees who recused themselves from voting on controversies for which they were conflicted. To evaluate the potential effect of noncompliance, we assumed that every vote cast by an ineligible conference attendee was cast in direction of the majority vote.

Results: Sixty-three panelists voted in at least one controversy at the final conference; the percentage of conflicted panelists varied from 6% to 39% for eight controversies. The compliance with the COI policy was 14 of 14 (100%) for one controversy, and varied from one of 19 (5%) to one of three (33%) in the remaining seven. In two of the eight controversies ("Compression device plus aspirin vs. low-molecular-weight heparin in thromboprophylaxis in orthopedic surgery" and "Low-molecular-weight heparin vs vitamin K antagonists for treatment"), the low compliance may have affected the final recommendations.

Conclusions: The low compliance raises concerns about implementation of COI restrictions in the context of anonymous voting.

Background

Conflicts of interest (COIs) are ubiquitous among guideline developers. Studies exploring COIs in specialty guidelines have found that panelists often (50%-90%) report ties with the pharmaceutical industry.[1-8] Nonfinancial COI (including intellectual COI) may also influence recommendations.[9,10]

Merely acknowledging potential COI may not reduce their impact on guideline recommendations. Excluding conflicted panelists altogether may compromise the quality of the guidelines and threaten their feasibility and, possibly, their acceptability. New strategies that limit the impact of COI while allowing the participation of clinical experts are needed.

The executive committee of the Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (AT9) developed a novel strategy designed to use experts' experience while limiting the potential impact of their COI on recommendations.[11] According to this strategy, conflicted clinical experts had input in identifying, summarizing, and interpreting the evidence. They were, however, excluded from the deliberations and voting that ultimately determined the direction and strength of recommendations with which they had conflicts (Table 1).

Some AT9 panels failed to achieve consensus on all recommendations. To achieve resolution, these remaining controversial issues were presented at the AT9 final conference, which was attended by panelists from across the entire guideline. Electronic voting on the controversial issues was conducted anonymously using an audience electronic response system. Conference attendees received instructions not to vote on issues for which they had an intellectual or financial COI (not prior to each voting, but repeatedly, and forcefully, throughout the conference). The number of people voting on some recommendations, however, appeared to exceed that which would be expected based on COIs. This led us to question the extent to which conference attendees complied with the COI policy in the anonymous electronic votes. The objective of this study is to explore the compliance of the AT9 final conference attendees with the voting aspect of the policy for managing COI when anonymously voting on the controversies.

Materials and Methods

Sixty-seven panelists from across the entire guideline attended the AT9 final conference (held February 10-11, 2011, Atlanta, Georgia). Anonymous electronic voting was conducted for nine controversies among the 628 total recommendations in the guidelines (Table 2). The ACCP staff kept record of the number of votes cast (in person or online) and of the result of voting. For each one of the controversies on which voting occurred during the final AT9 conference, we compared attendees' voting behavior with voting eligibility. The institutional review board of McMaster University approved the study (11-129-S).

Data Collection and Voting

We estimated the voting behavior using two sources:

1. A survey of the conference attendees (primary analysis): Two months after the conference, we conducted a survey (www.surveymonkey.com) of all the panelists who attended. We asked them whether they voted on each controversy and the reason for not voting when applicable (including not being present at that time, having a COI with the issue, not having enough understanding of the issue/evidence). The survey was not anonymous, but we deidentified the database to conduct the analysis.
2. The data recorded by the audience response system (secondary analysis).

We determined, for each attendee and each controversy on which there was a vote, whether a COI existed using four sources:

1. Self-report of conference attendees on the survey.
2. Public declarations of COI submitted to ACCP as part of the guideline process (which covered the 3 years preceding the launch of the AT9 development process: 2008-2010). It is important to note that these declarations included only the COI relevant to the AT9 chapter in which each panelist worked. However, during the final conference, attendees voted on controversies from several chapters.
3. Public declarations of COI in publications indexed in MEDLINE between 2008 and 2010: Teams of two investigators independently screened the full text of the publications available on MEDLINE between 2008 and 2010 for all conferences attendees. They extracted the COI statements of each panelist who attended the final conference and voted on at least one recommendation and judged the presence or absence of a COI with

each controversy. We considered as financial COI any declaration of payment, research funding, consultancy, advisory board membership, and the like from the manufacturer of one of the drugs or devices under discussion, which panelists had received in the 3 years prior to the conference. If the COI statement did not provide enough information to estimate the approximate date of the payment, we considered the COI relevant only if it was declared in an article published during 2010. We considered as intellectual COI any co-authorship on the trials directly bearing on the recommendation.

4. A review of the references cited in the AT9 guidelines: Two investigators independently screened all the references cited on the final version of AT9 relevant to the recommendations voted on at the conference. We used the same criteria previously described to judge the presence of financial or intellectual COI.

Disagreements were recorded and resolved by consensus. We assessed the agreement between the two reviewers with the unweighted κ statistic.

Data and Impact Analyses

The unit of analysis was the controversy. Using the four sources of vote eligibility mentioned earlier, for each controversy we determined the number of panelists in attendance free from COI (eligible for voting) and with financial or intellectual COI (ineligible for voting).

For primary analysis, using the data from the survey, we calculated compliance as the proportion of conflicted individuals who reported recusing themselves from voting due to COI on each controversy divided by the number of conflicted individuals in attendance

(we estimated the number of conflicted individuals in attendance using the four sources described in voting eligibility).

A small number of panelists answered the survey stating that they did not remember their voting behavior on some specific controversies. For the estimation of the compliance from survey data, we assumed that conflicted panelists actually voted on the specific controversies on which they did not remember their voting behavior and complement this approach by a sensitivity analysis assuming that they recused themselves from voting.

To complement the primary analysis, we secondarily estimated the compliance using the audience response system data. We first calculated the number of individuals who recused themselves from voting as the difference between the number of attendees and the number of observed votes. Because there were other reasons for not voting (eg, limiting understanding of the specific issue being discussed), we subtracted from this number the number of panelists that declared having recused themselves for reasons different to COI on the survey. By this procedure, we obtained an estimate of the number of individuals who recused themselves from voting, presumably to comply with the COI policy. To estimate the compliance, we divided our estimate of the number of individuals who complied with the COI policy by the number of conflicted individuals in attendance.

To explore the potential impact of the lack of compliance with the COI policy on the resolution of the controversies, we assumed that every vote cast by an ineligible conference attendee was cast in a direction to the majority vote (worst-case scenario).

Results

Sixty-seven panelists attended the final conference, and all responded to the survey. According to the self-report, 63 panelists voted in at least one controversy (there were two observers not allowed to vote, and two panelists felt that the issues being discussed when they were present were beyond their fields of expertise).

Voting Behavior and Eligibility

The number of votes in each controversy ranged from 39 to 57 according to the self-report on the survey and from 36 to 56 according to the ACCP audience response system. The difference between the number of self-reported and recorded votes was small (median difference, three votes; range, from one to six votes).

For the controversy regarding use of specialized anticoagulant clinics, 14 panelists declared recusing themselves due to COI when responding to the survey. For the remaining controversies, the proportion of panelists who declared recusing themselves due to COI was generally small (zero to four).

We identified 1,533 publications on MEDLINE between 2008 and 2010 in which at least one of the 63 panelists who voted in at least one controversy was a coauthor. We reviewed the 1,438 records available as full text and identified a median of 14 panelists with financial COI on the nine controversies on which votes occurred (range, from zero to 19). We also identified two instances of intellectual COI (co-authorship of a trial directly bearing on the recommendation). On two controversies ("Best practice suggestion on a structured clinic approach" and "Platelet count monitoring"), we did not find any relevant COI in published papers. The agreement between the reviewers on the

judgment about the presence or absence of COI was excellent (κ 0.82 for MEDLINE search, and 0.85 for hand search).

Compliance Analyses

We included in the analysis of the compliance the eight controversies in which we identified conflicted panelists in attendance (the controversy "Platelet count monitoring" was excluded).

Primary Analysis:

Using the information from the survey, the compliance with the restriction on voting varied from one of 19 (5%) to one of three (33%) on seven of the eight controversies. In one controversy ("Best practice suggestion on a structured clinic approach") we found 100% compliance (Table 3).

Sensitivity Analysis:

A median of one conflicted panelist (range, zero to three) did not remember whether they voted or not on specific controversies. The sensitivity analysis assuming that no votes occurred on these instances did not substantially change the results: The compliance varied from one of 15 (7%) to one of three (33%) on seven of the eight controversies and was 100% in the controversy "Best practice suggestion on a structured clinic approach."

Secondary Analysis:

Using the information from the audience response system, the compliance with the restriction on voting varied from one of 17 (6%) to eight of 17 (47%) on six of the eight controversies. On two controversies, the number of votes recorded was less than the number of eligible conference attendees. Because there are other potential reasons for not voting (eg, limited understanding of the specific issue), it was not possible to make inferences about the compliance in these controversies (Table 3).

Impact:

Applying the worst-case scenario approach, we found that in six of the eight controversies, the number of conflicted individuals who voted was not enough to potentially have changed the outcome. In the controversy "Intermittent pneumatic compression device (IPCD) + aspirin vs. low-molecular-weight heparins (LMWH) in the prevention of VTE in orthopedic surgical patients" (resolved in favor of LMWH, with a difference of nine votes in favor of LMWH), 18 conflicted panelist voted. Four of them had ties with the manufacturer of aspirin, five with the manufacturers of LMWH, and nine with both. Also, in the controversy "LMWH vs vitamin K antagonists (VKA) for treatment in patients without cancer" (resolved in favor of VKA, with a difference of 10 votes in favor of a weak recommendation for VKA), 15 conflicted panelists voted. Nine of them have ties with the manufacturers of LMWH, one with the manufacturer of warfarin, and five with both. The overall impact of these conflicts in the resolution of the controversies is uncertain.

Discussion

In this study, we evaluated compliance with the restriction on voting of the AT9 COI policy in eight controversies. We observed 100% compliance in one controversy and a relatively low compliance in the remaining seven (from 5% to 33% in the primary analysis and from 6% to 47% in the secondary analysis). Guideline panelists often did not acknowledge all their relevant COI and, despite being conflicted, voted in an anonymous system.

The low compliance observed in the AT9 final conference might reflect conscious or unconscious disagreements of conflicted individuals with the standards set by the COI policy and lack of monitoring of the voting system used.[12] Anonymous voting systems may offer the advantage of allowing participants to express their true opinion free from the implicit or explicit coercion of opinion leaders or strong advocates. This is a desirable objective because opinion leaders and strong advocates often have financial and intellectual COI. But the honor system method used during the AT9 final conference proved suboptimal because it created the circumstances that allowed individuals to break the rules. The novelty of the policy for managing COI may also have contributed to the low compliance. Although the rules for conflict were developed iteratively and stated in ways intended to be clear and explicit, we cannot rule out misunderstanding of the policies as a reason for low compliance.

Our study is limited by the source we used to identify COI that were apparently undeclared as part of the guideline development process. We relied primarily on public declarations in published papers, which usually provide little detail and focus almost exclusively on the financial ties of the authors with the manufacturers of drugs and

devices. Also, our strategy may have identified COI that were >3 years old, and, thus, were not relevant to AT9. On two controversies ("Best practice suggestion on a structured clinic approach" and "Platelet count monitoring"), we did not identify additional COI with our strategy. Given the absence of a clear relationship with drugs or devices in these controversies, our strategy may have underestimated the proportion of panelists with COI. This limitation may explain the high estimate of the compliance in the controversy "Best practice suggestion on a structured clinic approach," which is not concordant with the rest of the findings. An alternative explanation is the particularly vivid highlighting of COI (one attendee calling out to another, who was speaking at the time, "You are conflicted") during the discussion of the issue. Another potential limitation, which may impact the proportion of disclosed COI, was our strategy to obtain the disclosure of COI from conference attendees. The disclosures submitted to the American College of Chest Physicians (ACCP) included only the COI relevant to the specific chapter in which each panelist was involved, and did not include other chapters that were discussed at the final conference. Also, the focus of the survey was to establish the voting behavior: We asked conference attendees whether they voted or not on each controversy and, if not, what was the reason. We did not, however, ask attendees to specify all the relevant COI for each controversy. This limitation may have resulted in an apparent lower proportion of disclosed COI.

The estimation of compliance from the ACCP audience response system, although based on an objective count of votes, was limited by some uncertainty over who actually voted and who did not on particular recommendations. The estimation of the compliance from the survey is limited by possible selective memory of the respondents, but likely

represents a more accurate estimate of the true compliance because it is based on the voting behavior at an individual level.

The strengths of this study include the comprehensive search of additional COI and a survey response rate of 100%, which allowed us to know which panelists were in attendance at the time of the different controversies. Also, this is the first study addressing the compliance of individuals with COI (financial or intellectual) with a formal set of rules designed to allow expert input but limit the influence of COI.

How to limit the influence of COI on clinical practice guideline development while not excluding valuable expert input remains a challenge. The lack of compliance with the restriction on voting of the COI policy at the last conference could be viewed as a failure in terms of the process. The low compliance, however, had a small or no impact on the final product (ie, final recommendations) because of the small proportion of conflicted panelists in attendance. The ultimate standard for the success of the new policy is the extent to which it produced higher-quality guidelines: more consistent with the available evidence and the values and preferences of the individuals to whom the recommendations will be applied.

An alternative way to achieve compliance with the COI policy is through enforcement. We are developing a browser-based, interactive guideline development tool (www.guidelinedevelopment.org) that includes COI management and will prevent those who have declared a conflict during the guideline development process from voting. Our findings suggest that such a strategy, to be completely effective, would also require a search of prior publications. Future research should compare the benefits and limitations

of these and other strategies for managing COIs in clinical practice guidelines having the quality of the final product as the major outcome of interest.

Tables

Table 1 —Key Features of the New Policy for Managing COIs Implemented in the ACCP AT9

Key elements of the policy:

1. A methodologist free of financial or intellectual conflicts bears responsibility for each of the AT9 chapters.

2. There is equal emphasis on intellectual and financial COI.

3. Experts with many COI are excluded from the development of AT9.

4. Only panel members without conflicts can be involved in the development of the recommendation. Content experts with a conflict may contribute to preparing, summarizing, and interpreting the evidence but not to deliberations that ultimately determine the direction and strength of recommendations with which they are conflicted.

Definitions:

Financial COI: consultancies, advisory board membership, stock ownership and the like from companies whose products are the subject of a recommendation

Intellectual COI: Academic activities that create the potential for an attachment to a specific point of view that could unduly affect an individual's judgment about a specific recommendation. Such activities include authorship of original studies and peer-reviewed grant funding (government, not-for-profit organizations) directly bearing on a recommendation.

Table 2 —Controversies Decided by Electronic Anonymous Voting at the Final Conference of the ACCP AT9

Controversies
1. Should aspirin be included among the alternative therapies that the orthopedic chapter offers as a possible choice for people undergoing major orthopedic surgery compared to no antithrombotic therapy?
2. Should we actively recommend against aspirin or should it not be offered as a possible therapy for medical patients?
3. Should ASA be added as an alternative therapy for moderate and high risk non-orthopedic surgical patients compared to no antithrombotic therapy?
4. To make a best practice suggestion on a structured clinic approach instead of an evidence-based recommendation?
5. Intermittent Pneumatic Compression Device versus Low Molecular Weight Heparins in prevention of venous thromboembolism in orthopedic surgical patients
6. Intermittent Pneumatic Compression Device plus aspirin versus Low Molecular Weight Heparins in prevention of venous thromboembolism in orthopedic surgical patients
7. Should thrombolytic therapy vs. no thrombolytic therapy be recommended strongly or weakly?
8. Low Molecular Weight Heparins vs. Vitamin K antagonists for the long-term management of venous thromboembolism in patients without cancer
9. Should we recommend platelet count monitoring in different context of availability of sophisticated testing and heparin-induced thrombocytopenia incidence?

Table 3 —Summary of the Results

	Controversy							
	ASA as prophylaxis for orthopedic patients?	ASA as prophylaxis for medical patients?	ASA as prophylaxis for surgical patients?	Best practice suggestion on a structured clinic approach?	IPCD vs. LMWH in prophylaxis in orthopedic patients?	IPCD + ASA vs. LMWH in prophylaxis in orthopedic patients?	In PE, strong or weak recommendation of thrombolytic therapy?	LMWH vs. VKA for treatment in patients without cancer?
Number of votes	56	52	56	36	48	49	50	55
Eligible panelists	45	43	42	46	42	38	55	39
Recused for reasons different from COI	1	0	1	4	3	3	1	1
Conflicted panelists in attendance	15	17	16	14	15	19	3	18
Declared COI	1	3	2	14	1	1	1	4
COI identified in publications	14	14	14	0	15	19	2	14
Financial COI	14	14	14	0	14	19	1	14
Intellectual COI	0	0	0	0	1	0	1	0
Compliance (%) (from survey)	1/15 (7)	1/17 (6)	1/16 (6)	14/14 (100)	1/15 (7)	1/19 (5)	1/3 (33)	2/17 (11)
Compliance (%) (from ARS)	3/15 (20)	8/17 (47)	1/16 (6)	NA	7/15 (47)	6/19 (32)	NA	1/17 (6)
Impact of conflicted votes on the final decision	No	No	No	No	No	Potentially	No	Potentially

Declared COI: either in the survey or in the declaration submitted to the ACCP. ARS: Audience Response System NA: Not available. ASA: Acetylsalicylic Acid; IPCD: Intermittent Pneumatic Compression Device; LMWH: Low Molecular Weight Heparins; PE: Pulmonary Embolism; VKA: Vitamin K Antagonists; HIT: Heparin-induced thrombocytopenia.

References

1. Norris S, Holmer HK, Ogden LA, Burda BU. COI in clinical practice guideline development: a systematic review. PLoS One. 2011;6(10):e25153. Epub 2011 Oct 19.
2. Cosgrove L, Bursztajn HJ, Krimsky S, Anaya M, Walker J. Conflicts of interest and disclosure in the American Psychiatric Association's Clinical Practice Guidelines. Psychother Psychosom. 2009;78(4):228-32.
3. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011 Oct 11;343:d5621. doi: 10.1136/bmj.d5621.
4. Mendelson TB, Meltzer M, Campbell EG, Caplan AL, Kirkpatrick JN. Conflicts of interest in cardiovascular clinical practice guidelines. Arch Intern Med. 2011 Mar 28;171(6):577-84.
5. Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA. 2002 Feb 6;287(5):612-7.
6. Detsky AS. Sources of bias for authors of clinical practice guidelines. CMAJ. 2006 Oct 24;175(9):1033, 1035.
7. Johnson K. Medical societies scramble to include disclosure on clinical guidelines. CMAJ. 2011 May 17;183(8):E443-4.

8. Kahn JM, Rubenfeld GD. Financial disclosures in clinical practice guidelines. Crit Care Med. 2010 Aug;38(8):1755-6

9. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.

10. Norris S, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012 Jul;65(7):725-33.

11. Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, Lewis SZ, Nathanson I, Jaeschke R, Schünemann H. The vexing problem of guidelines and COI: a potential solution. Ann Intern Med. 2010 Jun 1;152(11):738-41

12. Baumeister RF, Heatherton TF. Self-Regulation Failure: An Overview. Psychological Inquiry. 1996:7(1):1-15

Authors' contributions:

Dr Neumann had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Dr Neumann: contributed to the conception and design of the study; collection, analysis, and interpretation of the data; and drafting of the article.

Dr Akl: contributed to the conception and design of the study, analysis and interpretation of the data, and drafting of the article.

Ms Valdes: contributed to data collection and analysis and drafting of the article.

Dr Bravo: contributed to data collection and analysis and drafting of the article

Ms Araos: contributed to data collection and analysis and drafting of the article.

Dr Kairouz: contributed to data collection and analysis and drafting of the article.

Dr Schünemann: contributed to the analysis and interpretation of the data and drafting of the article.

Dr Guyatt: contributed to the conception and design of the study, analysis and interpretation of the data, and drafting of the article.

Chapter 6: Discussion

Guidance in how to assess and use guidelines recommendations

The optimal use of clinical practice guidelines recommendations requires a judgment of its trustworthiness. In 2011, the United State Institute of Medicine (IOM) issued a report outlining several criteria to make this judgment.[1] The approach we proposed in the chapter "How to Use a Patient Management Recommendation: Clinical Practice Guidelines and Decision Analyses" has important similarities: we highlighted the importance of a clear and transparent process to develop recommendations, minimizing conflict of interests, using updated systematic reviews as source of evidence, offering clear explanations of the recommendation rationale while specifying both the confidence in the effect estimates and the strength of the recommendations and providing clear and actionable recommendations (Table 1).

However, there are some differences between the two approaches. We did not include a criterion regarding the composition of the guideline panel. Two reasons motivated this decision: 1. We believe that for clinicians (the target population) it would be difficult to

judge how appropriate were the panelists with the little information that is usually available in guideline documents; 2. The rest of the criteria are relatively hard to fulfill without an advanced knowledge of guideline methodology. Hence, it is highly unlikely to achieve the level of sophistication required without a knowledgeable panel.

Additionally, in our approach we highlighted three aspects that are not adequately covered by the IOM criteria: 1. Trustworthy recommendations should consider all the patient important outcomes and not a subset of them; 2. The strength of the recommendation has to match the underlying evidence regarding the desirable and undesirable consequences of the intervention; and 3. Weak recommendations should offer the information required to engage in shared decision-making. Recommendations that omit outcomes that are important to patients or are inadequately classified as strong may be misleading, since they may promote a course of action that does not capture adequately patients' values and preferences. On the other hand, weak recommendations that do not provide the information to engage in shared decision making leave clinicians without the tools to effectively implement the recommendation. We believe that these three criteria considerably expand the IOM work, and since they are simple to evaluate, they do not add a considerable burden to clinicians assessing recommendations.

We propose to approach recommendations according to their strength: strong recommendations can be applied to patients with a minimal discussion with the patient and without a detailed review of the supporting information while weak recommendations generally require a shared decision making approach and a detailed review of the underlying evidence. It is possible to argue that strong recommendations also may benefit from a shared decision making approach. In a context of unlimited time

and resources, we would agree. However, in reality clinicians face serious time and resource constrains and implementing detailed shared decision-making for every decision does not seem feasible. We promote a shared decision-making approach for when it is needed most: situations where there is no clearly superior alternative, and hence the appropriate course of action is highly dependent of the values and preferences of patients and caregivers.

Management of conflicts of interest

Traditionally, clinical practice guidelines have been developed by content experts who typically have both ties with the pharmaceutical industry and attachment to certain practices or opinions.[2-5] The strategy implemented at the American College of Chest Physicians 9th edition of the Antithrombotic Guidelines (AT9) to manage conflict of interests is novel in several ways: 1. It allows some participation of conflicted individuals; 2. It recognizes the existence of intellectual conflict of interests; and 3. It puts an unconflicted methodologist in charge of the recommendations.

It is relatively clear that simply acknowledging conflict of interests may not be enough to limit their influence. Thus, a common approach has been excluding panelists with conflict of interests. However, this is not always possible and in some circumstances it might be even problematic, since it might compromise the credibility and acceptability of the guideline. The strategy of allowing some participation but excluding conflicted panelists from critical deliberations for which they have conflicted interest seems a good compromise between the approaches of acknowledgement only and excluding participants.

Unfortunately, as we showed in the studies "Experiences with a novel policy for managing conflicts of interest of guideline developers: A descriptive qualitative study" and "Low anonymous voting compliance with the novel policy for managing conflicts of interest implemented in the ninth version of the American College of Chest Physicians antithrombotic guidelines" this strategy proved to be difficult to implement and was not completely successful. One of the challenges was to effectively exclude conflicted panelists from the critical deliberations, since: 1. Conflicted panelists in general did not recuse themselves from participating in such instances, and 2. The discussion about the direction and strength of the recommendations was a fluid process without a clear "final deliberation."

We found that the introduction of intellectual conflict of interests was one of the more controversial aspects of the new policy, and even at the end of the AT9 process not all the participants agreed with the idea. Attachment to certain practices or opinions is a natural product of academic activity, and henceforth, intellectual conflicts of interests are very common. The definition used in AT9 (being an author of one of the included trials) was very limited; if a more comprehensive definition of intellectual conflict of interest would had been used (like for example, any previous research or public opinion regarding the topic under discussion) many more panelists, including some of the unconflicted methodologists, should have been excluded from the critical deliberations. Given how ubiquitous are intellectual conflicts of interest, perhaps it is not possible to exclude panelists on this basis, and a better alternative might be to have panels with individuals representing different perspectives and hence different sets of intellectual conflict of interests.

Finally, the most successful change of the AT9 policy was the introduction of unconflicted methodologists at the leading role. Even the leading content experts, who saw a reduction of their influence and power within the panel, were at ease with this change. As it was highlighted by the IOM report,[1] methodologists are a vital component of any guideline panel. However, when they are in a position of power and can effectively influence the process, in addition to ensure an optimal use of the methods to develop the recommendations, they may help to control the influence of conflicts of interests.

Conclusions

In this thesis we offered explicit and easy-to-use guidance to clinicians regarding how to use guideline recommendations in the context of real life practice. This guide will appear in the upcoming third edition of the Journal of American Medical Association "Users' guide to medical literature", a book that is often used as reference text in many of the evidence based practice courses and workshops around the world.

We also expanded the knowledge of how to manage conflict of interests in guideline developers. Through the two studies evaluating the AT9 policy, we were able to show what aspects of the policy were successful and what aspects need to be reformulated. Clinical practice guidelines could be an efficient tool for evidence-based practice. There is an ethical mandate to make them as trustworthy as possible, which includes limiting the influence of conflict of interests.

Tables

Table 1 – US Institute of Medicine criteria compared to our criteria for assessing guidelines recommendation

US Institute of Medicine criteria[1]	Our approach
To be trustworthy, clinical guidelines should:	***Assessing recommendations:***
Be based on an explicit and transparent process	Is the Clinical Question Clear and Comprehensive?
Minimize the influence of conflict of interests	Were the Recommendations Based on the Current Best Evidence?
Be developed by a knowledgeable panel of methodologists and content experts	Are Values and Preferences Appropriately Specified for Each Outcome?
Be based on best current evidence, informed by systematic reviews	Do the Authors Indicate the Strength of Their Recommendations?
Explicitly consider the values and preferences of the people to whom the guidelines will be applied.	Is the Evidence Supporting the Recommendations Easily Understood?
Provide a clear explanation of reasoning underlying the recommendation, and provide rating of both the confidence in the effect estimates and the strength of the recommendations.	Was the Influence of Conflict of Interests Minimized?
Be articulated in a standardized form detailing precisely what the recommended action is and under what circumstances it should be performed	

References

1. Institute of Medicine. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, eds. Clinical practice guidelines we can trust. Washington (DC): National Academies Press; 2011. 2p.
2. Norris SL, Holmer HK, Ogden LA, Burda BU. Conflict of interest in clinical practice guideline development: a systematic review. PLoS One. 2011;6(10):e25153. Epub 2011 Oct 19.
3. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011 Oct 11;343:d5621
4. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005 Aug;2(8):e124. Epub 2005 Aug 30.
5. Norris SL, Burda BU, Holmer HK, Ogden LA, Fu R, Bero L, Schünemann H, Deyo R. Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. J Clin Epidemiol. 2012 Jul;65(7):725-33.