A guide for health professionals to interpret and use recommendations in guidelines developed with the GRADE approach

Ignacio Neumann[a,b,*], Nancy Santesso[a], Elie A. Akl[a,c], David M. Rind[d,e], Per Olav Vandvik[f], Pablo Alonso-Coello[g], Thomas Agoritsas[a,h], Reem A. Mustafa[a,i], Paul Elias Alexander[a], Holger Schünemann[a], Gordon H. Guyatt[a]

[a]Department of Clinical Epidemiology and Biostatistics, McMaster University, Room 2C12, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada
[b]Department of Internal Medicine, School of Medicine, Pontificia Universidad Catolica de Chile, Alameda 340, Santiago, Chile
[c]Department of Internal Medicine, American University of Beirut, American University of Beirut Medical Center, P.O. Box: 11-0236, Riad-El-Solh, Beirut 1107 2020, Lebanon
[d]Harvard Medical School, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA 02215, USA
[e]UpToDate, 230 Third Avenue, Waltham, MA 02451, USA
[f]Norwegian Knowledge Centre for the Health Services, PO Box 7004 St. Olavs plass, N-0130 Oslo, Norway
[g]Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau-CIBERESP), Sant Antoni Maria Claret, 167, Pabellón 18, 08025, Barcelona, Spain
[h]Division of Clinical Epidemiology & Division of General Internal Medicine, University Hospitals of Geneva, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva, Switzerland
[i]Department of Medicine, University of Missouri-Kansas City, Kansas City, 5100 Rockhill Rdm, MO, USA

Abstract

An increasing number of organizations worldwide are using new and improved standards for developing trustworthy clinical guidelines. One of such approaches, developed by the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group, offers systematic and transparent guidance in moving from evidence to recommendations. The GRADE strategy concentrates on four factors: the balance between benefits and harms, the certainty of the evidence, values and preferences, and resource considerations. However, it also considers issues around feasibility, equity, and acceptability of recommendations. GRADE distinguishes two types of recommendations: strong and weak. Strong recommendations reflect a clear preference for one alternative and should apply to all or almost all patients, obviating the need for a careful review of the evidence with each patient. Weak recommendations are appropriate when there is a close balance between desirable and undesirable consequences of alternative management strategies, uncertainty regarding the effects of the alternatives, uncertainty or variability in patients' values and preferences, or questionable cost-effectiveness. Weak recommendations usually require accessing the underlying evidence and a shared decision-making approach. Clinicians using GRADE recommendations should understand the meaning of the strength of the recommendation, be able to critically appraise the recommendation, and apply trustworthy recommendations according to their strength. © 2016 Elsevier Inc. All rights reserved.

1. Introduction

Trustworthy clinical practice guidelines aim to provide useful recommendations for the practice of evidence-based diagnosis and treatment at the point of care [1]. In recent years, the guideline community has seen major advances in the methods for developing clinical practice guidelines, including rigorous standards for development, and tools for assessing the methodological rigor and transparency [2,3].

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group offers a

What is new?

GRADE has become an international standard, adopted by more than 80 organizations worldwide, including the World Health Organization (WHO), the Cochrane Collaboration, the National Institute for Health and Care Excellence, the Scottish Intercollegiate Guideline Network, UpToDate, and Clinical Evidence. In this article, we present a users' guide for GRADE for health professionals including the interpretation of GRADE recommendations, their critical assessment, and their use in patient care.

Table 1. Outline

How should we interpret grade recommendations?
What does the strength mean?
What does the certainty of the evidence mean?
How should we critically assess grade recommendations?
Were all the relevant outcomes important to patients explicitly considered?
Was the recommendation based on the best current evidence?
Is the strength of the recommendation appropriate?
Is the recommendation clear and actionable?
Does the recommendation provide the necessary additional information?
How should we use grade recommendations?
Strong recommendations
Weak recommendations

systematic and transparent approach to summarize evidence, rate certainty of the evidence (also known as: confidence in or quality of evidence), and move from evidence to recommendations [4]. GRADE also distinguishes between recommendations that should apply to all or almost all patients, obviating the need for a careful review of the evidence with each patient, and recommendations that require accessing the underlying evidence and a shared decision-making approach.

GRADE has become an international standard, adopted by more than 80 organizations worldwide, including the World Health Organization (WHO), the Cochrane Collaboration, the National Institute for Health and Care Excellence, the Scottish Intercollegiate Guideline Network, UpToDate, and Clinical Evidence (see http://www.gradeworkinggroup.org/society/index.htm). In this article, we present a users' guide for GRADE for health professionals including the interpretation of GRADE recommendations, their critical assessment, and their use in patient care (Table 1). First, however, we will provide a rationale for seeking an evidence-based recommendation.

2. Why look for an evidence-based guideline recommendation?

Box 1 presents a woman with atrial fibrillation with an intermediate risk of a thromboembolic complication. Which of the following antithrombotic therapies is more appropriate: aspirin, other antiplatelet agents such as clopidogrel, a combination of aspirin plus other antiaggregants, warfarin, or new anticoagulants such as direct thrombin inhibitors or oral factor Xa inhibitors? To fully address the question, clinicians would need to integrate the information of several systematic reviews covering all the relevant comparisons and outcomes.

Although systematic reviews can provide evidence summaries reporting the estimates of the benefits and harms of the interventions, they do not integrate these factors with patients' values and preferences or resource considerations

to provide a suggested course of action. GRADE recommendations build on systematic reviews of randomized controlled trials and of observational studies but move beyond the evidence to integrate all the relevant factors into the formulation of recommendations. Clinicians, who typically have limited time to answer their clinical questions, are likely to prefer guidance rather than interpreting evidence summaries themselves.

3. How should we interpret a GRADE recommendation?

Box 1 provides two GRADE recommendations, which can apply to our clinical scenario. When using GRADE recommendations, clinicians need to conceptually understand the strength of the recommendation, the rating of the certainty of the evidence, and be able to access key additional information such as remarks, absolute effects of treatment alternatives, or considerations regarding values and preferences and cost.

3.1. What does the strength of recommendation mean?

Recommendations developed with the GRADE approach are classified as strong or weak. Strong recommendations reflect the guideline panel's high confidence that desirable consequences of the proposed course of action clearly outweigh the undesirable consequences or vice versa. They are usually framed as "we recommend" or "clinicians should."

Weak recommendations (also called conditional or discretionary where we suggest that those applying GRADE choose the terms that best convey the intended recommendation to the target audience), on the other hand, reflect the guideline panel's judgment that there is either a close balance between benefits and down sides (including adverse effects and burden of treatment), uncertainty regarding the magnitude of benefits and down sides, uncertainty or great variability in patients' values and preferences, or that the cost or burden of the proposed

---

Box 1 Clinical scenario: A 55-year-old patient with atrial fibrillation and hypertension. Which antithrombotic therapy is most appropriate?

You are a general internist following a 55-year-old woman with hypertension and nonvalvular atrial fibrillation. She has no history of diabetes, congestive heart failure, stroke, or renal failure. After discussion of the merits of rate vs. rhythm control and the relevant evidence, you and your patient have opted for the latter strategy with use of a beta blocker to control her ventricular rate and treat her hypertension, and a vitamin K antagonist (VKA) to prevent thromboembolic complications (CHADS2 score of 1). During the current visit, the patient inquires about dabigatran, "the new blood thinner that would save me from getting blood tests every couple of weeks." Indeed, the patient's INR has been challenging to maintain within target range in spite of close monitoring and adjustments.

Looking for a GRADE recommendation: You search for the most trustworthy information resource and find a recently published guideline [5] that includes the following set of recommendations that seem to answer your patient's question:

Should anticoagulation rather than no therapy be used in patients with atrial fibrillation?

Recommendation 1: For patients with atrial fibrillation who are at intermediate risk of stroke (eg, CHADS2 score 1), we recommend oral anticoagulation rather than no therapy (strong recommendation/moderate certainty of the evidence).

Should dabigatran 150 mg bid rather than VKA be used in patients with atrial fibrillation?

Recommendation 2: We suggest dabigatran 150 mg twice daily rather than adjusted-dose VKA (weak recommendation/moderate certainty of the evidence).

Remarks: Dabigatran is excreted primarily by the kidney. It has not been studied and is contraindicated in patients with severe renal impairment. Clinicians should be aware that there is no antidote for dabigatran. There are no long-term safety data.

---

intervention may not be justified. Weak recommendations are usually framed as "we suggest" or "we conditionally recommend" or "clinicians may."

3.2. What does the certainty of the evidence mean?

GRADE is applied to bodies of evidence addressing the relative merits of alternative management strategies. These bodies of evidence may be sparse (restricted to case reports or even verbal reports of clinicians) or substantial (ie, multiple randomized trials with large numbers of patients). The products of the GRADE approach to bodies of evidence are estimates of intervention effects for each of the patient-important outcomes, ideally obtained (if the evidence permits) through a meta-analysis.

The concept of certainty of the evidence represents the extent to which the effect estimates are sufficiently credible to support a particular recommendation. GRADE specifies four levels of certainty: high, moderate, low, and very low. This rating is determined for each relevant outcome by a systematic and transparent assessment of the study design, limitations of the body of evidence, and special circumstances that can increase our confidence. All relevant outcomes receive a certainty rating, with the overall rating of the evidence coming from the lowest across all the outcomes considered critical to decision making.

4. How should we critically assess grade recommendations?

Recommendations developed according to the minimal criteria for using GRADE will likely meet most key criteria for trustworthy guidelines (Table 2). A detailed description of all the trustworthiness criteria is beyond the scope of this article. We do, however, provide a guide to assessing whether the GRADE approach was appropriately used (Box 2).

4.1. Did the guideline panel explicitly consider all the relevant outcomes important to patients?

The balance between the benefits and the harms of the interventions will depend on what outcomes are considered.

Table 2. Trustworthiness criteria for clinical practice guidelines[a]

To be trustworthy, clinical guidelines should:
1. Be based on an explicit and transparent process
2. Minimize the influence of conflict of interests
3. Be developed by a knowledgeable panel of methodologists and content experts
4. Be based on best current evidence, informed by systematic reviews
5. Explicitly consider the values and preferences of the people to whom the guidelines will be applied
6. Provide a clear explanation of reasoning underlying the recommendation and provide rating of both the certainty of the evidence and the strength of the recommendations
7. Be articulated in a standardized form detailing precisely what the recommended action is and under what circumstances it should be performed

[a] Adherence to GRADE will ensure criteria 1, 4, 5, 6, and 7 are met and make it likely that criterion 3 is met.

Box 2 Using the guide

We have understood the concepts of strength of the recommendation and certainty of the evidence. The strong recommendation means that we can recommend anticoagulation without the need of a detailed discussion of the underlying evidence with our patient. However, we also know that we should individualize the choice of dabigatran vs. warfarin (weak recommendation). How can we appraise the trustworthiness of the identified guideline and use the latter weak GRADE recommendation in our clinical practice?

How should we critically assess grade recommendations?

Were all the relevant outcomes important to patients explicitly considered?

As we can see in Table 3, AT9 panelists explicitly considered the following outcomes: death, strokes, major bleeding, systemic embolism, and burden of treatment. It seems plausible that this selection includes all the outcomes that are important to patients and, at the same time, relevant to the decision of choosing dabigatran or warfarin in patients with atrial fibrillation who are at intermediate risk of stroke.

Was the recommendation based on the best current evidence?

In the methods section of the AT9 guidelines, we can 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" [7]. This strategy ensured that the recommendation for dabigatran was based on the best current evidence.

Is the strength of the recommendation appropriate?

The overall certainty of the evidence was rated as moderate for this clinical question. By looking at the absolute effect estimates in Table 3, we can see that although dabigatran offers a small reduction of strokes and reduces burden of treatment compared to VKA, the balance of benefits and harms is close. Further, in the recommendation's remarks, panelists pointed out that in contrast to warfarin, there is no antidote for dabigatran and limited data on its use in clinical practice. Finally, in the text that accompanies the recommendation, guideline panelists commented that it is uncertain whether well-controlled patients participating in home monitoring of VKA therapy would still prefer dabigatran

(variability in patients' values and preferences). Therefore, in this case, the close balance between benefits and harms and the potential variability of patients' values justify a weak recommendation.

Is the recommendation clear, actionable, and does it provide the necessary information?

The recommendation presented in Box 1 is clear and actionable: the target population is described in very specific terms and the suggested treatment (dabigatran) and the alternative (warfarin) are clearly specified. The recommendation also provides the rating of the overall certainty of the evidence and the grading of the recommendation strength. Finally, it is accompanied by a remarks section and the Table 3, which provides the anticipated absolute effects that we can use for shared decision making.

How should we use grade recommendations?

In summary, we deem this GRADE recommendation to be trustworthy and to provide all the necessary information. Now, how can we use this weak recommendation in our practice?

As weak recommendations are generally sensitive to patients' preferences, clinicians should consider a shared decision-making approach to tailor the application of the recommendation to the patient. Clinicians may start discussing the evidence from the recommendation's accompanying table (Table 3) and explaining that the use of dabigatran in comparison to VKA results in a small reduction of strokes (3 fewer strokes per 1,000 patient treated for a year) and a significant reduction in the burden of treatment (no need of lifestyle limitations, dietary restrictions, or frequent blood testing). At this point, we may summarize and say that dabigatran may also reduce mortality but may increase bleeding events. Specific remarks are also important to communicate to patients. We could explain that dabigatran is a relatively new drug with no antidote and no information about long-term adverse effects. Patients that consider VKA therapy burdensome or put a high value in the small reduction in strokes may choose dabigatran. However, patients comfortable with VKA therapy or patients who put a high value on avoiding potential long-term adverse events may choose VKA. Although this last option seems to be contrary to the recommendation, it is in essence not: a weak recommendation should be individualized to the specific patient values and preferences and to the specific circumstances.

Clinicians should judge whether the guideline panel included all the outcomes important to patients. Outcomes typically considered as patient important include mortality,

Table 3. Summary-of-findings table: should dabigatran 150 mg bid rather than vitamin K antagonists (VKAs) be used in patients with atrial fibrillation and intermediate risk of stroke (CHADS score = 1)

		Anticipated absolute effects, 1 year time frame		
Outcome	Relative effect (95% CI)	With VKA	With dabigatran	GRADE certainty of the evidence
Death	RR 0.89 (0.79–1.01)	38 per 1,000	4 fewer deaths per 1,000 (from 8 fewer to 0 more)	⊕⊕⊕○; moderate due to imprecision
Nonfatal stroke	RR 0.67 (0.52–0.86)	8 per 1,000	3 fewer strokes per 1,000 (from 1 to 4 fewer)	⊕⊕⊕○; moderate due to imprecision
Major bleeding	RR 1.07 (0.91–1.26)	13 per 1,000	1 more bleed per 1,000 (from 1 fewer to 3 more)	⊕⊕⊕○; moderate due to imprecision
Systemic embolism	RR 0.85 (0.39–1.84)	2 per 1,000	0 fewer embolism per 1,000 (from 1 fewer to 2 more)	⊕⊕⊕○; moderate due to imprecision
Burden of treatment	—	One pill daily, lifestyle modifications, frequent laboratory monitoring, dietary restrictions	Two pills daily	⊕⊕⊕⊕; high

Abbreviations: CI, confidence interval; GRADE, Grading of Recommendations Assessment, Development, and Evaluation; RR, relative risk.
Adapted from reference number [6].

morbidity (eg, major bleeding, acute exacerbation of a chronic disease, hospital admission), and patient-reported outcomes (eg, quality of life, functional status).

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

Surrogate outcomes (eg, lipid levels, bone density, cognitive function tests) have a variable link to patient-important outcomes but are never important in and of themselves.

The Antithrombotic Guidelines of the American College of Chest Physicians provide an illustration of the issue of surrogate outcomes. The 8th edition of the guidelines suggested international normalized ratio (INR) monitoring at an interval of no longer than every 4 weeks in patients treated with Vitamin K Antagonists (weak recommendation/low certainty of the evidence) [8] . 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 (AT9), however, suggested an INR testing frequency of up to 12 weeks rather than every 4 weeks (weak recommendation/moderate certainty of the evidence) [9]. 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. The outcomes were however, surrogate in the first case and—more appropriately—patient important in the second.

4.2. Was the recommendation based on the best current evidence?

Guideline panels should base their estimates of the benefits and harms of the intervention and their evaluation of the associated certainty of the evidence in systematic reviews. In the absence of current and well-conducted published systematic reviews, some guideline panels may conduct their own reviews or provide less systematic evidence summaries. Panels should, in the methods section of their guidelines, provide a description of the process used to identify and summarize the relevant evidence. Clinicians should judge to what extent this process is credible. Criteria to judge the credibility of systematic reviews—including explicit eligibility criteria, comprehensiveness of the search, and evaluation of risk of bias of primary studies—can be found elsewhere [10].

Recommendations that do not use the best current evidence risk promoting suboptimal or even harmful care. For example, for several years, guideline panels had ignored a substantial body of evidence suggesting the effectiveness of prophylaxis with quinolones in patients with postchemotherapy neutropenia [11] until the Infectious Diseases Society of America suggested the prophylactic use of antibiotics in its 2010 guideline [12]. This highlights the necessity for rapid updating of guidelines in areas under active investigation.

4.3. Is the strength of the recommendation appropriate?

As we will discuss later, the strength of the recommendation has important implications to effectively implement GRADE recommendations. Recommendations that are inappropriately graded as "strong" may provide a misleading message to clinicians.

When the overall certainty of the evidence is high or moderate, clinicians can be confident that the evidence is credible and thus will support a strong recommendation if the desirable and undesirable consequences are not closely balanced, there is reasonable confidence and limited variability in patients' values and preferences, and the benefit of the proposed course of action justifies its cost.

When the overall certainty of the evidence is low or very low, there is a substantial uncertainty regarding the impact

of the proposed course of action and clinicians should therefore expect weak recommendations. Sometimes, however, guideline panels can appropriately offer strong recommendations despite low or very low certainty of the evidence. Table 4 presents five paradigmatic situations in which this can occur. Clinicians should examine carefully a strong recommendation based on low or very low certainty of the evidence. If it does not correspond to any of the paradigmatic situations described in Table 4, it is safe to consider 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 certainty of the evidence. Of these 121, only 35 (29%) were consistent with one of the paradigmatic situations presented in Table 4, and thus clearly appropriate [18]. Additionally, a survey of WHO guidelines that used the GRADE method between 2007 and 2012 showed that among 456 recommendations, there were 160 strong recommendations based on low or very low certainty of the evidence. Only 25 (16%) of these recommendations were consistent with one of the paradigmatic situations [19]. These results highlight the need for caution when facing strong recommendations based on low or very low certainty of the evidence.

One mistake that panels make is to overlook the variability in what informed patients may choose. This is particularly relevant in the context of low or very low certainty of the evidence because different attitudes toward uncertain benefits or uncertain harms have the potential to change the perceived net effect of the intervention. For example, a guideline panel made a strong recommendation in favor of increasing potassium intake to reduce cardiovascular risk (strong recommendation/low certainty of the evidence) [20]. The intervention has no known adverse effects and almost no additional cost; however, the benefits are uncertain. In this circumstance, some patients will be willing to change their diet for the possibility of an uncertain benefit, but others will very reasonably decline. Given the uncertainty and the variability of patients' preferences, a weak recommendation would have been more appropriate.

4.4. Is the recommendation clear and actionable?

GRADE recommendations should include a rating of the overall certainty of the evidence across outcomes and a grade of the strength of the recommendation. In addition, they should be articulated in a standardized form detailing precisely to whom the recommendation applies, the recommended action, under what circumstances it should be performed, and the alternative to which the intervention was compared. Reports of the available evidence alone or vague suggested courses of actions, such as "consider" or "attempt" are generally not helpful: they may be consistent with several interpretations. For example, a guideline for the Diagnosis and Treatment of Diabetic Foot Infections

recommended: "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 (strong recommendation/low certainty of the evidence)" [21]. This recommendation it is too vague to be effectively implemented: what the panel means by "attempt to provide" or "well-coordinated approach" is uncertain, and which specialties are included in the "variety of specialties" is unclear.

4.5. Does the recommendation provide the necessary additional information?

Finally, recommendations—in particular weak recommendations—should explicitly provide the key underlying information necessary to act on the recommendation. This enhances the transparency of the recommendation development process, informs clinical judgment, and facilitates shared decision making. Such information could be typically found in the "summary-of-findings" (SoFs) tables, which provide the certainty ratings for all the relevant outcomes and the numerical values for the relative and absolute estimates of the effect. Table 3 summarizes an SoF table relevant for the clinical scenario presented in Box 1. As we will discuss later, SoF tables can be used as a tool to engage in shared decision making (see Section 5.2). Currently, the GRADE working group, in the context of the group's DECIDE project (http://www.decide-collaboration.eu), is developing new presentation formats for recommendations and improved formats to provide clinicians supporting information summarizing the benefits and harms of interventions, the judgments regarding the certainty of the evidence, the typical patients' values and preferences, and cost considerations [22].

5. How should we use graded recommendations?

In this section, we will discuss how to use recommendations developed with GRADE, noting the critical difference between strong and weak recommendations.

5.1. Strong recommendations

If the panel has correctly distinguished between recommendations that warrant a designation of "strong" and those that do not, clinicians can apply strong recommendations to all or almost all the patients in all or almost all the circumstances without thorough (or even cursory) review of the underlying evidence and without a detailed discussion with the patient.

For example, the Allergic Rhinitis and its Impact on Asthma guideline recommended intranasal glucocorticoids rather than intranasal antihistamines for treatment of allergic rhinitis in adults (strong recommendation/high certainty of the evidence) [23]. This recommendation was based on an important reduction of symptoms with

Table 4. Five paradigmatic situations that justify strong recommendations based on low or very low certainty of the evidence

Paradigmatic situation	Certainty of the evidence 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 certainty of the evidence). 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 [13].
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 [14].
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 in less harm or cost for one of the competing	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-certainty evidence suggests that initial *Helicobacter pylori* eradication in patients with early-stage extranodal

(Continued)

Table 4. Continued

Paradigmatic situation	Certainty of the evidence for health outcomes (quality of evidence)		Balance of benefits and harms	Values and preferences	Resource considerations	Recommendation	Example
	Benefits	Harms					
			alternatives				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 [15].
High 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 certainty of the evidence suggests similar effects of different anticoagulants. However, indirect evidence (low certainty of the evidence) 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 [16].
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 men with androgen deficiency, testosterone supplementation likely improves

(Continued)

Table 4. Continued

Paradigmatic situation	Certainty of the evidence for health outcomes (quality of evidence)		Balance of benefits and harms	Values and preferences	Resource considerations	Recommendation	Example
	Benefits	Harms					
							quality of life. Low-certainty 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 [17].

glucocorticoids (rhinorrhea, nasal blockage, and itching) with no difference in 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 confidence intervals), 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 in most circumstances.

Another example of a strong recommendation, this time based on very low certainty of the evidence, is the recommendation of the WHO in favor of the treatment with oseltamivir in patients with suspected or confirmed Avian Influenza [13]. Indirect evidence from seasonal influenza provides some basis to infer that oseltamivir might be beneficial in the treatment of Avian Influenza. However, at the time of the WHO guideline publication, there was only evidence from a small case series about the effects of oseltamivir in those patients. On the other hand, Avian Influenza is a life-threatening disease, to which there is no other alternative treatment available. These circumstances match the first paradigmatic situation presented in Table 4. If clinicians agree with the panel's judgment, as in the previous example, they could adopt the practice of prescribing oseltamivir to patients with suspected or confirmed Avian Influenza without reviewing extensively the underlying evidence or a detailed discussion with the patient.

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.

5.2. Weak recommendations

With careful consideration of the evidence, as well as 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 over 600 recommendations issued in the American College of Chest Physicians 9th edition of the Antithrombotic guidelines (AT9) guidelines were weak.

Clinicians can apply weak recommendations to most patients, but not to all. To use weak recommendations, clinicians will need to consider the key factors (or conditions) driving the strength of the recommendation (eg, fine balance between benefit and harm) and to understand the underlying evidence. In addition, this type of recommendation is frequently sensitive to patients' preferences. Consequently, a shared decision-making approach, involving a discussion with the patient about the potential benefits and harms of the proposed course of action is typically the optimal way to ensure that the decision reflects both the best evidence available and the patient's values and preference.

For example, the AT9 guidelines suggested antepartum prophylaxis with low-molecular-weight heparin (LMWH) rather than clinical vigilance or routine care in pregnant women at moderate to high risk of thrombosis (weak recommendation/low certainty of the evidence) [16]. This is a weak recommendation because there was low certainty of the evidence (due to imprecision and the use of indirect evidence from other populations at high risk of thrombosis) and expected variability in patients' values and preferences regarding the relative importance of thrombosis, bleeding, and the burden of treatment. However, the panel judged that

desirable consequences of antepartum LMWH would probably outweigh undesirable consequences for most patients and therefore its use.

In such situations, clinicians should follow prescribed approaches to shared decision making [24–26]. The process starts with conveying to patients the awareness of choice, then exploring treatment alternatives in light of the relative values and preferences that patients assign to relevant outcomes. This presentation should include estimates of the absolute (rather than relative) impact of treatment alternatives and the certainty of the evidence. For example, in our antithrombotic prophylaxis example, clinicians may start by discussing with the patient the estimates for symptomatic thromboembolism with LMWH during pregnancy vs. no treatment: 51 fewer events per 1,000 patients. Then, the clinician could inform the patient that the use of LMWH may slightly increase the risk major maternal bleeds (up to 3 more events per 1,000 followed over the pregnancy) and mention the potential burden of treatment that daily injections for several months will represent (low certainty of the evidence for all outcomes aside from the burden of injections).

Typical patients might, despite the uncertainty, 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. Other patients, however, might think differently. For example, a study exploring women' preferences about antepartum use of LMWH found that most women at either low or high risk for thrombosis would prefer to use LMWH during pregnancy [27]. Given it is a weak recommendation, both courses of action (to use and to not use LMWH) would be consistent with the guideline panel's guidance.

Weak recommendations are also mandated in the context of moderate or high certainty of the evidence when there is a close tradeoff between benefits and harms or uncertainty or variability in patients' values and preferences. For example, the American College of Physicians suggested the use of cholinesterase inhibitors or memantine in patients with dementia (weak recommendation/moderate certainty of the evidence) [6]. This recommendation is based on evidence from randomized trials warranting moderate 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. As before, to use this recommendation, clinicians may present the estimates of benefits and harms of the intervention to the patients or their families. By doing this, clinicians can help patients make a decision consistent with both the best evidence available and the patients/families' values and preferences.

6. Conclusions

In this article, we provide a guide to interpret, critically appraise, and use GRADE recommendations from the clinical health professional perspective and provide examples of real recommendations developed with GRADE approach in different fields. The guide is designed to help clinicians to better understand the meaning and consequences of strong and weak recommendations in practice and to facilitate their application in real-life decision making.

References

[1] Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ 1999;318:527–30.
[2] Graham R, Mancher M, Dianne MW, Greenfeld S, Steinberg E. Clinical practice guidelines we can trust. 1st ed. Washington, D.C: Institute of Medicine of the National Academies; 2011.
[3] Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. J Clin Epidemiol 2010;63:1308–11.
[4] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94.
[5] You JJ, Singer DE, Howard PA, Lane DA, Eckman MH, Fang MC, et al. American College of Chest Physicians. 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;141:e531S–75S.
[6] Qaseem A, Snow V, Cross JT Jr, Forciea MA, Hopkins R Jr, Shekelle P, et al. 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;148:370–8.
[7] Guyatt GH, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, et al. American College of Chest Physicians. 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;141:53S–70S.
[8] 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;133:160S–98S.
[9] Holbrook A, Schulman S, Witt DM, Vandvik PO, Fish J, Kovacs MJ, et al, 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:e152S–84S.
[10] Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA 2014;312:171–9.

[11] Hughes WT, Armstrong D, Bodey GP, Bow EJ, Brown AE, Calandra T, et al. 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. Clin Infect Dis 2002;34:730—51.

[12] Freifeld AG, Bow EJ, Sepkowitz KA, Boeckh MJ, Ito JI, Mullen CA, et al, 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:e56—93.

[13] Schünemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, et al. 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;7:21—31.

[14] Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al, 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;183:788—824.

[15] Freedman AS, Lister A, Connor RF. Management of gastrointestinal lymphomas. In: Basow DS, editor. UpToDate. Waltham, MA: UpToDate; 2013.

[16] 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:e691S—736S.

[17] Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, et al. Task Force, Endocrine Society. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2010; 95:2536—59.

[18] 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;98:3246—52.

[19] Alexander PE, Brito JP, Neumann I, Gionfriddo MR, Bero L, Djulbegovic B, et al. World Health Organization strong recommendations based on low-quality evidence (study quality) are frequent and often inconsistent with GRADE guidance. J Clin Epidemiol 2016;72:98—106.

[20] WHO. Guideline. Potassium intake for adults and children. Geneva: World Health Organization (WHO); 2012:Lipsky BA, Berendt AR, Cornia PB, Pile JC, Peters EJ.

[21] Armstrong DG, Deery HG, Embil JM, Joseph WS, Karchmer AW, Pinzur MS, et al. 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:e132—73.

[22] Treweek S, Oxman AD, Alderson P, Bossuyt PM, Brandt L, Bro¿ek J, et al. DECIDE Consortium. Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence (DECIDE): protocol and preliminary results. Implement Sci 2013;8:6.

[23] Brozek JL, Bousquet J, Baena-Cagnani CE, Bonini S, Canonica GW, Casale TB, et al. Global Allergy and Asthma European Network, Grading of Recommendations Assessment, Development and Evaluation Working Group. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines: 2010 revision. J Allergy Clin Immunol 2010; 126(3):466—76.

[24] Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012;27:1361—7.

[25] Stiggelbout AM, Van der Weijden T, De Wit MP, Frosch D, Légaré F, Montori VM, et al. Shared decision making: really putting patients at the centre of healthcare. BMJ 2012;344:e256.

[26] van der Weijden T, Boivin A, Burgers J, Schünemann HJ, Elwyn G. Clinical practice guidelines and patient decision aids. An inevitable relationship. J Clin Epidemiol 2012;65:584—9.

[27] Bates SM, Alonso-Coello P, Tikkinen KAO, Ebrahim S, Lopes LC, McDonald SD, et al. Women's values and preferences and health state valuations for thromboprophylaxis during pregnancy: A cross-sectional interview study. Thromb Res 2016; http://dx.doi.org/10.1016/j.thromres.2015.12.015.