GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines

Pablo Alonso-Coello,[1,2] Andrew D Oxman,[3] Jenny Moberg,[3] Romina Brignardello-Petersen,[2,4] Elie A Akl,[2,5] Marina Davoli,[6] Shaun Treweek,[7] Reem A Mustafa,[2,8] Per O Vandvik,[3] Joerg Meerpohl,[9] Gordon H Guyatt,[2,10] Holger J Schünemann,[2,10] the GRADE Working Group

For numbered affiliations see end of article.

Correspondence to:
P Alonso-Coello
palonso@santpau.cat

Introduction

Clinicians regularly face situations with two or more alternative actions. Each alternative often has different advantages and disadvantages, including differences in effectiveness, adverse effects, costs and other factors (criteria). To make these choices, clinicians rely on recommendations from clinical practice guidelines,[1] other recommendations (such as from colleagues or experts) or implicit rules for decision making, such as based on their personal experience or what others do. To ensure trustworthiness, clinical practice guidelines are made by groups of people (guideline panels) with relevant skills, perspectives, and knowledge; they are informed by the best available evidence; and they are systematically developed.[1-4]

In the first article in this series, we described GRADE Evidence to Decision (EtD) frameworks and their rationale for different types of decisions.[5] In this second article, we describe the use of EtD frameworks for clinical recommendations and how they can help clinicians and patients who use those recommendations.

We will use the scenario in box 1 to illustrate the use of EtD frameworks for clinical recommendations.[6-8] The question posed for the panel in this scenario was: "Should patients with atrial fibrillation and a moderate to high risk of stroke who are currently taking warfarin switch to dabigatran?" The panel specified the question details, including the population, intervention, comparison, and outcomes (PICO),[9] the setting for which the recommendation is intended, and the perspective they have taken (box 2).

The EtD framework for clinical recommendations was developed as part of the DECIDE project, using an iterative process.[10-12] The starting point for EtD frameworks for clinical recommendations was the GRADE Working Group's approach for going from evidence to clinical recommendations.[13-15] We further developed the EtD framework for clinical recommendations based on reviews of relevant literature and handbooks for clinical guidelines, brainstorming and discussion by the authors, feedback from stakeholders, user testing, and application of the framework to a range of recommendations in workshops and real guidelines. Detailed methods used to develop EtD frameworks are available in the DECIDE project protocol.[10] Appendix 1 is a glossary of terminology used in EtD frameworks.

The general structure of EtD frameworks is the same for clinical recommendations, coverage decisions, and health system or public health recommendations and decisions.[5] The EtD frameworks include three main sections: formulating the question, assessing the evidence and additional considerations for each criterion, and drawing conclusions. Appendix 2 is an example an EtD framework for the scenario in box 1. In this article we focus on elements of EtD frameworks for clinical recommendations that are of particular relevance to clinicians and patients.

Formulating the question

When formulating the relevant questions, panels should specify the patients, intervention, comparison, and outcomes (PICO),[9] their perspective, subgroups for which the evidence and their judgments and recommendation might differ from an overall recommendation, and the settings for which the recommendation is intended.

SUMMARY POINTS

- Clinicians do not have the time or resources to consider the underlying evidence for the myriad decisions they must make each day and, as a consequence, rely on recommendations from clinical practice guidelines
- Guideline panels should consider all the relevant factors (criteria) that influence a decision or recommendation in a structured, explicit, and transparent way and provide clinicians with clear and actionable recommendations
- The GRADE working group has developed Evidence to Decision (EtD) frameworks for different types of decisions and recommendations.
- The purpose of the Evidence to Decision (EtD) frameworks is to help groups of people (panels) use evidence in a structured and transparent way to inform decisions in the context of clinical recommendations, coverage decisions, and health system or public health recommendations and decisions. In this article we will describe EtD frameworks for clinical practice recommendations
- The general structure of the EtD framework for clinical recommendations is similar to EtD frameworks for other types of recommendations and decisions, and includes formulation of the question, an assessment of the different criteria, and conclusions
- Clinical recommendations require considering criteria differently, depending on whether an individual patient or a population perspective is taken. For example, from an individual patient's perspective, out-of-pocket costs are an important consideration, whereas, from a population perspective, resource use (not only out-of-pocket costs) and cost effectiveness are important
- From a population perspective, equity, acceptability, and feasibility are also important considerations, whereas the importance of these criteria is often limited from an individual patient perspective
- Specific subgroups for which different recommendations may be required should be clearly identified and considered in relation to each criterion because judgments might vary across subgroups

---

Box 1: Clinical scenario

Warfarin reduces the risk for ischaemic stroke in patients with atrial fibrillation, but increases the risk for haemorrhage and requires frequent blood tests and clinic visits to monitor the international normalised ratio (INR) and adjust the dose. Apixaban, dabigatran, and rivaroxaban are new, fixed-dose, oral anticoagulants, each of which has been compared with warfarin in randomised trials.[6-8]

Dabigatran is a direct thrombin inhibitor. The RE-LY (Randomized Evaluation of Long-Term Anticoagulation Therapy) trial was an international, multicentre, randomised trial in which 18 113 patients with atrial fibrillation at increased risk for stroke (CHADS2 score ≥1) were randomly assigned to receive low dose dabigatran (110 mg twice daily), high dose dabigatran (150 mg twice daily), or adjusted dose warfarin.[6] The median follow-up was two years. Outcomes were better with the higher dose of dabigatran.

A guideline panel from a national health system guideline programme is faced with the question: "Should dabigatran or warfarin be used for atrial fibrillation in patients with a moderate to high risk of stroke?"

Box 2: Evidence to Decision (EtD) framework—Question formulation section*

Should patients with atrial fibrillation and a moderate to high risk of stroke who are currently taking warfarin switch to dabigatran?

Problem: Patients with atrial fibrillation and a moderate to high risk of stroke taking warfarin

Intervention: Dabigatran (150 mg) daily

Comparison: Warfarin

Main outcomes: Death, stroke, major bleeding, myocardial infarction, treatment burden

Setting: High resource setting

Perspective: Health system

Subgroups: Patients who are well controlled with warfarin

Background: Warfarin reduces the risk for ischaemic stroke in patients with atrial fibrillation but increases the risk for haemorrhage and requires frequent blood tests and clinic visits to monitor the international normalised ratio (INR) and adjust the dose. Apixaban, dabigatran, and rivaroxaban are new, fixed-dose, oral anticoagulants, each of which has been compared with warfarin in randomised trials.[6-8]

Dabigatran is a direct thrombin inhibitor. The RE-LY (Randomized Evaluation of Long-Term Anticoagulation Therapy) trial was an international, multicentre, randomised trial in which 18 113 patients with atrial fibrillation at increased risk for stroke (CHADS2 score ≥1) were randomly assigned to receive low dose dabigatran (110 mg twice daily), high dose dabigatran (150 mg twice daily), or adjusted dose warfarin.[6] The median follow-up was two years. Outcomes were better with the higher dose of dabigatran.

*Templates used for EtD frameworks are adapted for specific types of decisions. The one shown here is for a clinical recommendation from a population perspective.

Guideline panels are often not explicit about the perspective they are taking. This can lead to confusion and, sometimes, to inappropriate recommendations. For example, from an individual patient perspective, whether the net desirable effect of an intervention, such as dabigatran, is worth the out-of-pocket costs can be critical for making a decision. This can be an issue if the government or insurance does not pay the full cost of the drug or if coverage is restricted. Total resource requirements (outside of out-of-pocket costs), cost effectiveness (from a population perspective), and impacts on equity are unlikely to be critical for personal choices. However, from a population perspective, such as the one taken by the National Institute for Health and Care Excellence (NICE), decisions affect how limited healthcare resources are used. Total resource requirements, cost effectiveness (from a broader perspective than that of individual patients),

equity, acceptability, and feasibility considerations can drive a recommendation.

Consideration of different perspectives can lead to formulation of different recommendations. For example, taking a population perspective, a panel might recommend restricted use of new anticoagulants for patients with atrial fibrillation because of their large costs and a small increase in desirable health effects for patients who are already well controlled with warfarin (the comparator). In contrast, taking an individual perspective in the context of small out-of-pocket costs (for patients with insurance that pays most or all of the cost of the new anticoagulants), a panel might recommend new anticoagulants because they are less burdensome than warfarin, which requires daily medication, lifestyle limitations, dietary restrictions, and frequent blood tests and clinic visits.

The remit of the organisation making a recommendation usually determines the specific perspective that a panel takes. For example, a national guideline developer, such as NICE, might take the perspective of the government or the department of health, given its mandate to ensure optimal use of the health budget in the country. A professional society, on the other hand, might take an individual patient perspective with a view towards providing guidance to individual patients and clinicians making individual patient choices.

Recommendations can differ across subgroups of the population originally considered when formulating the question. This may be due to differences in people (such as differences in baseline risk as assessed by the CHADS$_2$ score), differences in interventions (such as different doses or different drugs within the same class), differences in comparisons (such as different levels of international normalised ratio (INR) control with warfarin), or different settings (such as differences in access to a thrombosis clinic).[16] For the question in box 2, the panel paid particular attention to patients with good INR control. The rationale for this is that patients taking warfarin with good INR control have better outcomes than patients with poor INR control, and, consequently, the desirable health effects of dabigatran compared with warfarin are less.[16]

Assessing the criteria considered

EtD frameworks for clinical recommendations from a population perspective include 12 criteria. For recommendations from an individual patient perspective, some criteria differ in how they are applied (table 1). The technical team or panel may use research evidence from systematic reviews or single studies to inform judgments about the effects of the intervention and other criteria. For example, they might use an epidemiological study of the baseline risk for an outcome in the setting(s) of interest, a systematic review or, occasionally, a single study of the effects of an intervention when that is all that is available, a systematic review or a single study of how much people value the relevant outcomes, or an economic analysis.

The source of the evidence summarised in the framework should be referenced, and any limitations of how

Table 1 | Criteria for clinical recommendations from a population and an individual patient perspective

Population perspective	Individual patient perspective
Is the problem a priority (from a population perspective)?	Is the problem a priority (from the perspective of individual patients)?
How substantial are the desirable anticipated effects?	
How substantial are the undesirable anticipated effects?	
What is the overall certainty of the evidence of effects?	
Is there important uncertainty about or variability in how much people value the main outcomes?	
Does the balance between desirable and undesirable effects favour the intervention or the comparison?	
How large are the resource requirements (costs)?	Does the cost effectiveness of the intervention (the out-of-pocket cost relative to the net desirable effect) favour the intervention or the comparison?
What is the certainty of the evidence of resource requirements (costs)?	
Does the cost effectiveness of the intervention favour the intervention or the comparison?	
What would be the impact on health equity?	
Is the intervention acceptable to key stakeholders?	Is the intervention acceptable to patients, their care givers, and healthcare providers?
Is the intervention feasible to implement?	Is the intervention feasible for patients, their care givers, and healthcare providers?

the evidence was summarised should be noted, particularly when the source is not a systematic review. If the technical team does not find any evidence for a criterion, they should note the lack of evidence and include any relevant information or assumptions used to make a judgment under "Additional considerations."

Technical teams or panels can reduce the burden of preparing frameworks. For example, in a guideline it is often the case that individual systematic reviews answer one or more criteria across several EtD frameworks. They can duplicate frameworks electronically,[5] avoiding the need to replicate work. In other instances, it can be clear from early in the evidence synthesis that it is unnecessary to review the evidence for some criteria. For example, if there is high certainty evidence for large harms and small benefits, the direction and strength of a recommendation might be clear, making it unnecessary to review the evidence for the rest of the criteria. This can also work in the opposite direction. For example, a panel charged with making recommendations for funding of health technologies recently considered faecal microbiota transplantation for recurrent *Clostridium difficile* infection. The overwhelming evidence of large benefit for this extremely serious, costly condition with minimal adverse effects made detailed consideration of other criteria unnecessary. Under such circumstances, panels can then rapidly consider the rest of the criteria without systematically reviewing the evidence and provide their rationale under "Additional considerations."

Additional considerations can include:

- Other evidence, such as estimates from routinely collected data
- Plausible consequences for which no evidence was found (such as logical reasons for anticipating a potential reduction in inequities) or plausible reasons for anticipating that the intervention (option) might not be acceptable to key stakeholders or might be difficult to implement
- Any assumptions that were made and, if relevant, the basis for those assumptions
- Explanations of the basis for a judgment, if a judgment does not flow directly from the research evidence (such as the logic underlying a judgment

about the balance between desirable and undesirable effects)
- Documenting voting results or relevant discussions by the panel.

Is the problem a priority?
The more serious a problem is, the more likely it is that an intervention that addresses the problem should be a priority or should be recommended (if it is effective with minimal harms or burden). For example, from a population perspective, helpful interventions for conditions that are fatal or disabling are likely to be a higher priority, and to be recommended, than the ones for transient conditions or those that cause only minor and reversible distress. Panels might decide that all of the problems that a particular guideline addresses are equally important as part of a prioritisation process, making this criterion irrelevant. They might also argue that, from an individual patient perspective, the importance of the problem is not relevant (if the patient has a condition and wants to do something about it, it will always be a priority). For instance, a panel considered the use of ultrasound scanning as a complement to mammography in women at high risk of breast cancer (1% of women) in whom magnetic resonance imaging (the complementary imaging procedure of choice) was contraindicated (a very small proportion of high risk women). Although this affects only a very small group of women, the issue is highly relevant to that population.

However, the importance of a problem can sometimes affect decisions made by individual patients. For example, patients' priorities for primary prevention might affect the strength of recommendations, as some problems (risks) might be more important than others, or patients' baseline risk might be so low that prevention would not be a priority, even if it was effective. Similarly, patients with comorbidities and their carers might need to consider a number of different treatments and might need to prioritise these based on how important the problems are.

How substantial are the desirable and undesirable anticipated effects?
Summaries of findings, such as fig 1, provide estimates of the effects of the interventions being compared on

Outcome	Plain language summary	Absolute Effect		Relative effect (95% CI) Nº of participants & studies	Certainty of the evidence (GRADE)
		Without Dabigatran	With dabigatran		
Death Follow-up: 2 years	*Dabigatran probably reduces the risk of dying.*	76 per 1000	68 per 1000	RR 0.89 (0.79 to 1.01) Based on data from 12098 patients in 1 study	⊕⊕⊕◯ Moderate
		Difference 8 less per 1000 patients (95% CI: 17 less to 1 more per 1000 patients)			
Nonfatal stroke Follow-up: 2 years	*Dabigatran reduces the risk of a nonfatal stroke.*	34 per 1000	23 per 1000	RR 0.64 (0.51 to 0.81) Based on data from 12098 patients in 1 study	⊕⊕⊕⊕ High
		Difference 11 less per 1000 patients (95% CI: 15 to 6 less per 1000 patients)			
Major bleeding Follow-up: 2 years	*There probably is little or no difference in the risk of major nonfatal bleeding outside the brain.*	70 per 1000	66 per 1000	RR 0.94 (0.83 to 1.08) Based on data from 12098 patients in 1 study	⊕⊕⊕◯ Moderate
		Difference 4 less per 1000 patients (95% CI: 12 less to 6 more per 1000 patients)			
Myocardial infarction Follow-up: 2 years	*Dabigatran probably increases the risk of a heart attack.*	12 per 1000	16 per 1000	RR 1.38 (1.00 to 1.91) Based on data from 12098 patients in 1 study	⊕⊕⊕◯ Moderate
		Difference 4 more per 1000 patients (95% CI: 0 to 9 more per 1000 patients)			
Burden of treatment Follow-up: 2 years	*Dabigatran reduces the burden of treatment.*	1000 per 1000	0 per 1000	RR 0.00 (0.00 to 0.00) Based on data from 12098 patients in 1 study	⊕⊕⊕⊕ High
		Difference 1000 less per 1000 patients (95% CI: 1000 to 1000 less per 1000 patients)			

Fig 1 | "Summary of findings" table: dabigatran versus warfarin for atrial fibrillation.[4] An interactive version of this table is at http://isof.epistemonikos.org/#finding/5377108ff30d0c7233205f13

the outcomes of interest. While this summary is based on a systematic review that identified a single large study, a typical summary of findings is based on systematic reviews of multiple studies.[17]

The more substantial the desirable effects, the more likely it is that an intervention should be recommended. Conversely, the more substantial the undesirable effects (including the relative burden of interventions), the less likely it is that an intervention should be recommended. Judgments about how substantial effects are should take into account the absolute magnitude of the effect (such as the proportion of people who would benefit) and the importance of the outcome (such as how much it is valued by the people affected).

What is the overall certainty (quality) of the evidence of effects?
The less certain the evidence is for the main outcomes (desirable and undesirable effects, including the burden), the less likely it is that a strong recommendation (appendix 1) should be made for an intervention, and the more likely it is that the intervention should be evaluated, if implemented.[18 19]

In the scenario in box 1, the overall certainty of the evidence (the lowest certainty for the outcomes that are critical for a decision) is moderate, primarily because of risk of bias (table 2). The certainty of the evidence for the effect in well controlled patients was considered low because of imprecision in addition to risk of bias

(fig 2). Factors that should be assessed when evaluating the certainty of the evidence for each outcome include risk of bias, inconsistency, indirectness, imprecision, and publication bias.[18 19] Although the certainty of the evidence was low for this subgroup of patients, the panel judged that the subgroup effect was credible[20]— that is, that the subgroup estimates provided a better basis for decision making than the overall estimates.

Is there important uncertainty about or variability in how much people value the main outcomes?
Typically, people place a higher value on avoiding a stroke than on avoiding serious gastrointestinal bleeding (box 3). Uncertainty about how much those affected (patients or their carers) value the outcomes of interest can be a reason to make a weak (conditional) rather than a strong recommendation (appendix 1). Variability in how patients value the main outcomes (to the extent that individuals with different values would make different decisions) is another reason for a weak recommendation. For example, some patients might place a lower value on avoiding a stroke compared with avoiding serious gastrointestinal bleeding or the burden of warfarin treatment than other patients.

A systematic review found that there is moderate certainty of the evidence that typical patients place approximately three times more value on avoiding a stroke than on avoiding major gastrointestinal

Table 2 | Evidence profile: dabigatran 150 mg daily versus warfarin for atrial fibrillation[6]

	Evidence assessment						Summary of findings			Anticipated absolute effects	
No of participants (studies) Follow-up period	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty (quality) of the evidence	Study event rates (%) With warfarin	With dabigatran	Relative risk (95% CI)	Risk with warfarin	Risk difference with dabigatran (95% CI)
Death (critical outcome: assessed with No of people died)											
12 098 (1 study) 2 years	Serious*	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	Moderate due to risk of bias*	487/6022 (7.6%)†	438/6076 (7.2%)	0.9 (0.79 to 1.01)	81 per 1000†	8 fewer per 1000 (17 fewer to 1 more)
All cause stroke and systemic embolism (critical outcome; assessed with No of patients with stroke and systemic embolism)											
12 098 (1 study) 2 years	Serious*	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	Moderate due to risk of bias*	202/6022 (3.4%)†	134/6076 (2.2%)	0.66 (0.54 to 0.82)	34 per 1000†	11 fewer per 1000 (6 to 15 fewer)
Major bleeding (critical outcome; assessed with No of patients with major bleed)											
12 098 (1 study) 2 years	Serious*	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	Moderate due to risk of bias*	421/6022 (7%)†	399/6076 (6.6%)	0.94 (0.83 to 1.08)	70 per 1000†	4 fewer per 1000 (12 fewer to 6 more)
Myocardial infarction (critical outcome; assessed with No of patients with myocardial infarction)											
12 098 (1 study) 2 years	Serious*	No serious inconsistency	No serious indirectness	Serious‡	Undetected	Low due to risk of bias and imprecision*‡	75/6022 (1.2%)†	97/6076 (1.6%)	1.29 (0.96 to 1.73)	12 per 1000†	4 more per 1000 (0 to 9 more)
Burden of treatment (critical outcome)											
Expert opinion	—	—	—	—	—	High§	1000/1000	0/1000			

*Control (warfarin) arm was unblinded. Downgraded by from high to moderate quality of the evidence due to open label warfarin arm, which has potential to introduce performance bias to make dabigatran look better. (In the apixaban and rivaroxaban trials patients and healthcare professionals were blinded in both arms.)
†Average risk from RE-LY trial.
‡Further downgraded from moderate to low quality evidence because of imprecision. There were fewer than 300 events and a wide confidence interval.
§Although there is uncertainty about the number of clinic visits required with dabigatran, it is certain that fewer are required than for warfarin and warfarin has additional burdens of lifestyle limitations, dietary restrictions, and frequent blood testing.

bleeding,[21] and moderate variability in values across patients. In the dabigatran scenario it is uncertain whether patients who are well controlled with warfarin would value avoiding the burden of taking warfarin more than the possible downsides of switching to dabigatran. There is probably important variability in how much value patients place on avoiding the burden of warfarin; that is, it is likely that some patients would choose to switch and others would choose not to switch based on the burden of taking warfarin. Similarly, 40-50 year old women might make different choices about breast cancer screening because of differences in how averse they are to the undesirable effects and burden of screening mammography.[22 23]

Does the balance between the desirable and undesirable effects favour the intervention or the comparison?

Judgments about the balance between the desirable and undesirable effects need to take into account the preceding four judgments (the magnitude of the desirable and undesirable effects, the certainty of the evidence of effects, and how much those affected value the outcomes). In the scenario in box 1, the panel decided that the balance of desirable and undesirable effects probably favours dabigatran (fig 1 and table 2). However, the balance is less clear for the subgroup of patients who are well controlled with warfarin. These patients would have similar outcomes, apart from the greater burden of taking warfarin compared with dabigatran (fig 2).[6 16] For this subgroup, panels might be more inclined to judge the balance as probably favouring warfarin or not favouring either option. Uncertainty about potential adverse effects of dabigatran and compliance with taking the drug could increase uncertainty about the balance between the desirable and undesirable effects (appendix 2).

How large are the resource requirements (costs)?

The greater the cost, the less likely it is that an intervention will be recommended. If resource use is considered to be critical to a decision about a recommendation the more likely it is that resource use should be formally evaluated. Which costs and savings are included depend on the perspective that is taken.

In the dabigatran example, for an estimated 66 000 patients with atrial fibrillation, dabigatran was estimated to cost €30 million per year more than warfarin, considering both the cost of the drugs and clinic visits for monitoring. The difference in the estimated lifetime cost of the two drugs, for a population of 66 000 patients, was €308 million more for dabigatran.[24] From an individual perspective the costs depend on how much the target population is likely to pay out-of-pocket. This might be all the cost (if dabigatran is not covered by insurance), a proportion (if it is partially covered) or nothing (if it is fully covered).

5

Outcome	Plain language summary	Absolute Effect		Relative effect (95% CI) Nº of participants & studies	Certainty of the evidence (GRADE)
		Without Dabigatran	With dabigatran		
Death Follow-up: 2 years	*Dabigatran may increase the risk of dying.*	60 per 1000	65 per 1000	RR 1.08 (0.81 to 1.44) Based on data from 3023 patients in 1 study	⊕⊕◯◯ Low
		Difference 5 more per 1000 patients (95% CI: 11 less to 26 more per 1000 patients)			
Nonfatal stroke Follow-up: 2 years	*Dabigatran may reduce the risk of a nonfatal stroke.*	27 per 1000	25 per 1000	RR 0.95 (0.61 to 1.48) Based on data from 3023 patients in 1 study	⊕⊕◯◯ Low
		Difference 2 less per 1000 patients (95% CI: 11 less to 13 more per 1000 patients)			
Major bleeding Follow-up: 2 years	*Dabigatran may increase the risk of major bleeding.*	62 per 1000	71 per 1000	RR 1.16 (0.88 to 1.54) Based on data from 3023 patients in 1 study	⊕⊕◯◯ Low
		Difference 9 more per 1000 patients (95% CI: 7 less to 33 more per 1000 patients)			
Myocardial infarction Follow-up: 2 years	*Dabigatran probably increases the risk of a heart attack.*	12 per 1000	16 per 1000	RR 1.38 (1.00 to 1.91) Based on data from 12098 patients in 1 study	⊕⊕⊕◯ Moderate
		Difference 4 more per 1000 patients (95% CI: 0 to 9 more per 1000 patients)			
Burden of treatment Follow-up: 2 years	*Dabigatran reduces the burden of treatment.*	1000 per 1000	0 per 1000	RR 0.00 (0.00 to 0.00) Based on data from 12098 patients in 1 study	⊕⊕⊕⊕ High
		Difference 1000 less per 1000 patients (95% CI: 1000 to 1000 less per 1000 patients)			

Fig 2 | Summary of findings: patients who are well controlled taking warfarin.[14] An interactive version of this table is at http://isof.epistemonikos.org/#finding/537730b3f30d0c7233205f14

What is the certainty (quality) of the evidence of resource requirements (costs)?

If resource use is considered to be critical for a recommendation, the less certain the evidence is for resource requirements, the less likely it is that a panel should make a strong recommendation for or against an intervention. Judgments about the certainty of the evidence for resource requirements are similar to judgments about the evidence of effects.[25] In the scenario in box 1, there is important uncertainty about resource use for dabigatran and other new anticoagulants (appendix 2).

Does the cost effectiveness of the intervention favour the intervention or the comparison?

The greater the cost in relation to the net benefit, the less likely it is that an intervention should be recommended. Judgments about the cost effectiveness of an intervention need to take into account several criteria, including

- The balance between the desirable and undesirable effects (the net benefit), the certainty of the evidence of effects, and uncertainty about or variability in how much people value the main outcomes
- Resource requirements (costs) and uncertainty about the costs.

Several economic evaluations have assessed the cost effectiveness of dabigatran for stroke prevention in atrial fibrillation patients in different settings.[26] These models generally found dabigatran to be cost effective, but the incremental cost effectiveness ratios (ICERs) varied considerably between them (appendix 2). In a Norwegian economic evaluation, the ICER was less than a suggested threshold of €70 000 per quality adjusted life year in 80% of simulations. It was not cost effective for the subgroup of patients who are well

---

Box 3: Expressing the importance (value) of outcomes as a measure of utility (or disutility)

- One way of expressing the value of a health state is to use utility values, a measure the strength of the preference people have for a specific health state, from zero (for death) to one (for perfect health)
- A disutility is a reduction in utility. For example, a severe stroke might have a utility value of 0.10, which is a reduction or disutility of 0.90 compared with being healthy, a minor stroke might have a utility value of 0.75 (a disutility of 0.25), and a serious gastrointestinal bleed a utility value of 0.90 (a disutility of 0.10)
- Such values indicate that the relative importance of a severe stroke (or how much people value avoiding a severe stroke) is more than that of a minor stroke, which is more than that of a gastrointestinal bleed
- Evidence about utilities can come from studies that have measured utility values or, ideally, from systematic reviews of those studies. Evidence can also come from studies that directly measure the choices people make when presented with the probabilities of the desirable and undesirable effects, a description of those outcomes (health states), and information about when they would occur and how long they would last. Qualitative research evidence can also sometimes inform judgments about how much people value different outcomes

Box 4: Justification and detailed justification of a strong and a conditional recommendation

Strong recommendation

The guideline panel **recommends** in favour of oral anticoagulation rather than aspirin for patients with atrial fibrillation who are at high risk of stroke (such as CHADS$_2$ score ≥2).

Justification—In patients with atrial fibrillation with a high risk of stroke (such as CHADS$_2$ score ≥2) the reduction in the number of strokes outweighs the increase in the number of additional non-fatal major extracranial bleeds and the burden of oral anticoagulation, and oral anticoagulation is cost effective. Therefore, the panel made a strong recommendation in favour of oral anticoagulation rather than aspirin for patients with a high risk of stroke.

Conditional recommendation

For patients with atrial fibrillation who have a moderate to high risk of stroke (CHADS$_2$ score ≥1) the guideline panel **suggests** they should switch to dabigatran only if they are not well controlled with warfarin, despite good adherence (conditional recommendation, moderate certainty of the evidence).

Justification—Overall, and particularly for patients who are not well controlled with warfarin, the balance of desirable and undesirable effects favours dabigatran. However, the panel made a weak (conditional) recommendation in favour of switching to dabigatran for patients who are not well controlled with warfarin because of concerns about the cost of dabigatran (from a population perspective) and uncertainty about the balance of the desirable and undesirable effects (including uncertainty about the risk of rare severe adverse effects). The panel made a weak (conditional) recommendation against switching to dabigatran for patients who are well controlled with warfarin because there may be little or no reduction in the risk of strokes and warfarin is cost effective compared with dabigatran for these patients, but warfarin treatment might be very burdensome for some patients.

controlled with warfarin, or overall for thresholds below €38 000 (2012 Euros).[24]

In addition, if a cost effectiveness ratio from a formal economic evaluation is used, panels also should consider how robust the estimate is when single or multiple variables in the model are varied (one-way and multi-variable sensitivity analyses), whether the economic evaluation is reliable, and if a published economic evaluation was used, how applicable it is for the setting(s) of interest.

What would be the impact on health equities?

Interventions that reduce inequities are more likely to be recommended than ones that do not (or ones that increase inequities).[27 28] In the dabigatran example, the panel considered that dabigatran might reduce inequities for people who do not have easy access to INR testing, although there was no direct research evidence for this.

Is the intervention acceptable to key stakeholders?

The less acceptable an intervention is to key stakeholders (including patients), the less likely it is that it should be recommended, or if it is recommended, the more likely it is that an implementation strategy might be needed to address concerns about acceptability. An intervention might be unacceptable due to the distribution of the desirable and undesirable effects and costs; that is, who benefits (or who is harmed) and who pays (or saves). For example, people who would have increased costs or burdens without experiencing the benefits of an intervention might find this unacceptable.

Disagreement about ethical principles (such as autonomy, non-maleficence, beneficence, or justice) is another reason why some stakeholders might find an intervention unacceptable.[29 30] Some ethical considerations, such as autonomy, may be important enough to some organisations or panels that they might elect to consider these separately, either as a detailed judgment or as a criterion.

In the dabigatran example, some patients and clinicians might also be opposed to restrict its use. These are unlikely to be reasons not to restrict the use of dabigatran but might be important implementation considerations.

Is the intervention feasible to implement?

The less feasible (capable of being accomplished or brought about) an intervention is, the less likely it is that it should be recommended to clinicians. Barriers to implementing an intervention can also modify the strength of a recommendation. Clinicians might find it unhelpful to receive strong recommendations if the interventions are not implementable in their settings. However, if the target audience is policymakers, a panel might want to make a strong recommendation, despite barriers that currently make it difficult or impossible for clinicians to adhere to the recommendation. Panels can also incorporate consideration of critical barriers, such as the availability of the intervention, directly into their recommendations. More commonly, panels can assist those responsible for implementing recommendations by addressing key barriers to implementing their recommendation in their conclusions.[31]

Putting it all together

How important each of the above criteria is for a recommendation can vary. To make a recommendation, a panel must consider the implication and importance of each of the above judgments. In many cases, this will be straightforward and not require detailed consideration. However, when there is uncertainty or disagreement, it can help to explicitly consider this for each criterion.

Based on their overall assessment across criteria, panels must reach a conclusion about the direction of their recommendation (for or against the intervention) and the strength of their recommendation.[13 14] They should provide a justification for their recommendation, based on the criteria used in their assessment. They can reach these conclusions in different ways, including using informal or formal consensus processes or voting. For straightforward recommendations, informal consensus processes are often sufficient.

In the dabigatran example, the panel made a weak recommendation in favour of switching to dabigatran only for patients who are not well controlled with warfarin despite good adherence. Their concern about the cost of dabigatran (from a population perspective) and uncertainty about the balance of the desirable and undesirable effects (including uncertainty about the risk of rare severe adverse effects) was the reason for their making a weak recommendation. The panel made a weak (conditional) recommendation against

switching to dabigatran for patients who are well controlled with warfarin because there may be little or no reduction in the risk of strokes and warfarin is cost effective compared with dabigatran for these patients, but warfarin treatment might be very burdensome for some patients. A justification summarises the panel's judgments for each of the criteria that were most important for their decision (box 4).

In another scenario a panel made a strong recommendation in favour of oral anticoagulation rather than aspirin for patients with atrial fibrillation who are at high risk of stroke (such as $CHADS_2$ score ≥2). The panel was confident that the reduction in the number of strokes clearly outweighed the number of additional non-fatal major extracranial bleeds, and made a strong recommendation on this basis (box 4).

The panel's conclusions about implementation considerations should specify key concerns about the feasibility and acceptability of the intervention and strategies to address those concerns. For example, if the panel recommended limiting the use of dabigatran to patients who are not well controlled with warfarin, it might anticipate that this would be unacceptable to some key stakeholders, including the pharmaceutical company, some clinicians, and some patients. In this case, it might be important to be prepared to address objections raised by those stakeholders. In addition, it might be important to consider strategies to ensure that dabigatran is prescribed only to those patients.

Finally, panels can specify any indicators that should be monitored, when the recommendation is implemented, and priorities for further research to address important uncertainties (appendix 2).[32] For example, given that there is uncertainty about the costs and possible adverse effects of dabigatran, and potential concerns about adherence to the recommendation, it might be important to monitor and evaluate these.

Final remarks

EtD frameworks for clinical practice recommendations provide a structured and transparent approach for guideline panels. The framework helps ensure consideration of key criteria that determine whether an intervention should be recommended and that judgments are informed by the best available evidence. Frameworks are also a way for panels to make guideline users aware of the rationale (justification) for their recommendations.

Feedback from workshops and use of EtD frameworks by guideline panels has been uniformly positive. The most important concern that has been raised about EtD frameworks by guideline developers is that they are complex and require additional resources for preparation. Similar concerns have been raised about the complexity of other elements of the GRADE approach to making judgments about the certainty of evidence and strength of recommendations. It is, however, the judgments themselves that are complex, not the GRADE approach or EtD frameworks.

The challenge faced by GRADE or any other approach, is to keep the approach to making these judgments as simple as possible, but no simpler. While it might be possible not to consider some criteria, there is a risk of ignoring potentially important criteria that panels should take into account when making a recommendation. However, guideline panels need to make pragmatic decisions. For example, it is not always possible to undertake a full economic evaluation or to conduct systematic reviews for each criterion for which this might be relevant. Nor is it always necessary. Use of EtD frameworks does not require this, but they do require transparent consideration of which judgments are important for a recommendation and what evidence is used to inform each judgment.

AUTHOR AFFILIATIONS

[1]Iberoamerican Cochrane Centre, CIBERESP-IIB Sant Pau, Barcelona, Spain

[2]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada

[3]Global Health Unit, Norwegian Institute of Public Health, Oslo, Norway

[4]Evidence Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Santiago, Chile

[5]Department of Internal Medicine, Clinical Epidemiology Unit, American University of Beirut Medical Center, Beirut, Lebanon

[6]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy

[7]Health Services Research Unit, University of Aberdeen, Aberdeen, UK

[8]Department of Medicine, University of Missouri-Kansas City, Kansas City, Missouri, USA

[9]Cochrane Germany, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

[10]Department of Medicine, McMaster University, Hamilton, Canada

1   Institute of Medicine. *Clinical practice guidelines we can trust.* National Academies Press, 2011, p4. http://books.nap.edu/openbook.php?record_id=13058&page=4

2   Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. *Ann Intern Med* 2011;154:774-5. doi:10.7326/0003-4819-154-11-201106070-00011.

3   Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Board of Trustees of the Guidelines International Network. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;156:525-31. doi:10.7326/0003-4819-156-7-201204030-00009.

4   Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 2014;186:E123-42. doi:10.1503/cmaj.131237.

5   Alonso-Coello P, Schunemann H, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ* 2016;353:i2016. doi:10.1136/bmj.i2016.

6   Connolly SJ, Ezekowitz MD, Yusuf S, et al. RE-LY Steering Committee and Investigators. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139-51. doi:10.1056/NEJMoa0905561.

7   Patel MR, Mahaffey KW, Garg J, et al. ROCKET AF Investigators. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883-91. doi:10.1056/NEJMoa1009638.

8   Granger CB, Alexander JH, McMurray JJ, et al. ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981-92. doi:10.1056/NEJMoa1107039.

9   Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12-3.

10 Treweek S, Oxman AD, Alderson P, 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. doi:10.1186/1748-5908-8-6.

11 Rosenbaum SE. *Improving the user experience of evidence. A design approach to evidence-informed health care.* Oslo College of Architecture and Design, 2010: 108-16.

12 Morville P. User experience design. Sematic Studios LLC, 2004. www.semanticstudios.com/publications/semantics/000029.php

13 Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66:719-25. doi:10.1016/j.jclinepi.2012.03.013.

14 Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726-35. doi:10.1016/j.jclinepi.2013.02.003.

15 Guyatt GH, Oxman AD, Kunz R, et al. GRADE Working Group. Going from evidence to recommendations. *BMJ* 2008;336:1049-51. doi:10.1136/bmj.39493.646875.AE.

16 Wallentin L, Yusuf S, Ezekowitz MD, et al. RE-LY investigators. Efficacy and safety of dabigatran compared with warfarin at different levels of international normalised ratio control for stroke prevention in atrial fibrillation: an analysis of the RE-LY trial. *Lancet* 2010;376:975-83. doi:10.1016/S0140-6736(10)61194-4.

17 Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383-94. doi:10.1016/j.jclinepi.2010.04.026.

18 Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-6. doi:10.1016/j.jclinepi.2010.07.015.

19 Guyatt GH, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151-7. doi:10.1016/j.jclinepi.2012.01.006.

20 Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA* 2014;311:405-11. doi:10.1001/jama.2013.285063.

21 MacLean S, Mulla S, Akl EA, et al. 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(Suppl):e1S-23S. doi:10.1378/chest.11-2290.

22 Earle CC, Chapman RH, Baker CS, et al. Systematic overview of cost-utility assessments in oncology. *J Clin Oncol* 2000;18:3302-17.

23 Hall J, Gerard K, Salkeld G, Richardson J. A cost utility analysis of mammography screening in Australia. *Soc Sci Med* 1992;34:993-1004. doi:10.1016/0277-9536(92)90130-I.

24 Hamidi V, Wisløff T, Ringerike T, Linnestad KK, Harboe I, Klemp M. [Treatment of patients with acute stroke in stroke units (with or without early supported discharge)]. Rapport fra Kunnskapssenteret nr 18-2010. Oslo: Nasjonalt kunnskapssenter for helsetjenesten, 2010. www.kunnskapssenteret.no/Publikasjoner/_attachment/10097

25 Brunetti M, Shemilt I, Pregno S, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. *J Clin Epidemiol* 2013;66:140-50. doi:10.1016/j.jclinepi.2012.04.012.

26 Sorensen SV, Peng S, Monz BU, Bradley-Kennedy C, Kansal AR. A comparative analysis of models used to evaluate the cost-effectiveness of dabigatran versus warfarin for the prevention of stroke in atrial fibrillation. *Pharmacoeconomics* 2013;31:589-604. doi:10.1007/s40273-013-0035-8.

27 Dans AM, Dans LF, Oxman AD, et al. Addressing inequities in clinical practice guidelines. *Clin Epidemiol* 2007;60:540-6.doi:10.1016/j.jclinepi.2006.10.008.

28 Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 12. Incorporating considerations of equity. *Health Res Policy Syst* 2006;4:24. doi:10.1186/1478-4505-4-24.

29 Beauchamp TL, Childress JF. *Principles of biomedical ethics.*7th ed. Oxford University Press, 2013.

30 Saarni SI, Braunack-Mayer A, Hofmann B, van der Wilt GJ. Different methods for ethical analysis in health technology assessment: an empirical study. *Int J Technol Assess Health Care* 2011;27:305-12. doi:10.1017/S0266462311000444.

31 Flottorp SA, Oxman AD, Krause J, et al. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. *Implement Sci* 2013;8:35. doi:10.1186/1748-5908-8-35.

32 Fretheim A, Oxman AD, Lavis JN, Lewin S. SUPPORT tools for evidence-informed policymaking in health 18: Planning monitoring and evaluation of policies. *Health Res Policy Syst* 2009;7(Suppl 1):S18. doi:10.1186/1478-4505-7-S1-S18.

Appendix 1: Evidence to Decision frameworks: terminology

Appendix 2: GRADE Evidence to Decision framework for clinical recommendations

Appendix 3: GRADE Recommendation to Decision (RtD) presentation of an evidence to decision (EtD) framework for a clinical recommendation