GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction

Pablo Alonso-Coello,[1,2] Holger J Schünemann,[2,3] Jenny Moberg,[4] Romina Brignardello-Petersen,[2,5] Elie A Akl,[2,6] Marina Davoli,[7] Shaun Treweek,[8] Reem A Mustafa,[2,9] Gabriel Rada,[10,11,12] Sarah Rosenbaum,[4] Angela Morelli,[4] Gordon H Guyatt,[2,3] Andrew D Oxman[4] the GRADE Working Group

For numbered affiliations see end of article.

Correspondence to:
A D Oxman oxman@online.no

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;353:i2016
http://dx.doi.org/10.1136/bmj.i2016

Introduction

Healthcare decision making is complex. Decision-making processes and the factors (criteria) that decision makers should consider vary for different types of decisions, including clinical recommendations, coverage decisions, and health system or public health recommendations or decisions.[1-4] However, some criteria are relevant for all of these decisions, including the anticipated effects of the options being considered, the certainty of the evidence for those effects (also referred to as quality of evidence or confidence in effect estimates), and the costs and feasibility of the options. Decision makers must make judgments about each relevant factor, informed by the best evidence that is available to them.

Often, the processes that decision makers use, the criteria that they consider and the evidence that they use to reach their judgments are unclear.[5-8] They may omit important criteria, give undue weight to some criteria, or not use the best available evidence. Systematic and transparent systems for decision making can help to ensure that all important criteria are considered and that the best available research evidence informs decisions.

Clinicians depend on clinical practice guidelines. Rigorously developed guidelines synthesise the available relevant research, facilitating the translation of evidence into recommendations for clinical practice.[9] However, the quality of guidelines is often suboptimal.[10 11] If guidelines are not developed systematically and transparently, clinicians are not able to decide whether to rely on them or to explore disagreements when faced with conflicting recommendations.[12]

The GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group has previously developed and refined a system to assess the certainty of evidence of effects and strength of recommendations.[13-15] More than 100 organisations globally, including the World Health Organization, the Cochrane Collaboration, and the National Institute for Health and Care Excellence (NICE) now use or have adopted the principles of the GRADE system. Recently, through the DECIDE (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence) project (http://www.decide-collaboration.eu),[16] funded by the European Union, the GRADE Working Group has developed the Evidence to Decision (EtD) frameworks to support the process of moving from evidence to decisions. We have developed EtD frameworks for making clinical recommendations, coverage decisions, and health system or public health recommendations and decisions. The frameworks build on the GRADE approach to assessing the strength of recommendations.[17-19]

We developed EtD frameworks using an iterative process that is described in the project protocol.[16] The starting point for EtD frameworks was the GRADE Working Group's approach for moving from evidence to clinical recommendations.[17-19] We iteratively developed the frameworks based on reviews of relevant literature,[1-4] brainstorming, feedback from stakeholders,[20] application of EtD frameworks to a variety of recommendations and decisions, and user testing. We strove for consistency across EtD frameworks for different types of decisions, but, because of differences in the nature of the decisions, there are some differences in the frameworks. In appendix 1, we have provided a glossary of terms used in EtD frameworks, including certainty of the evidence, decisions, recommendations, and strength of recommendations.

This series of two articles describing the EtD frameworks is targeted at guideline developers and users of guidelines. This first article introduces the frameworks. It describes their purpose, development, and structure. It also describes how different organisations can adapt the frameworks to their own contexts and decision-making processes. The second article presents the framework for clinical recommendations.[21]

SUMMARY POINTS

- Clinicians, guideline developers, and policymakers sometimes neglect important criteria, give undue weight to criteria, and do not use the best available evidence to inform their judgments
- Explicit and transparent systems for decision making can help to ensure that all important criteria are considered and that decisions are informed by the best available research evidence
- The purpose of Evidence to Decision (EtD) frameworks is to help people 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
- EtD frameworks have a common structure that includes formulation of the question, an assessment of the evidence, and drawing conclusions, though there are some differences between frameworks for each type of decision
- EtD frameworks inform users about the judgments that were made and the evidence supporting those judgments by making the basis for decisions transparent to target audiences
- EtD frameworks also facilitate dissemination of recommendations and enable decision makers in other jurisdictions to adopt recommendations or decisions, or adapt them to their context

Purpose of the frameworks

The main purpose of the 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.

EtD frameworks:

- Facilitate adaptation of recommendations and decisions to specific contexts
- Inform panels about the relative pros and cons of the interventions or options being considered
- Ensure that panels consider important criteria for making a decision
- Provide panels with a concise summary of the best available evidence to inform their judgments about each criterion
- Help panels structure discussion and identify reasons for disagreements, making the process and the basis for decisions structured and transparent.

EtD frameworks assist users of recommendations by

- Enabling them to understand the judgments made by the panel and the evidence supporting those judgments
- Helping them to decide whether recommendations can and should be implemented in their own settings.

Box 1: Example for application of Evidence to Decision (EtD) framework

Use of bedaquiline to treat multidrug-resistant tuberculosis

"WHO estimates that up to half a million new cases of multidrug-resistant tuberculosis (MDR-TB) occur worldwide each year. Current treatment regimens for MDR-TB present many challenges: treatment lasts 20 months or more, requiring daily administration of drugs that are more toxic, less effective, and far more expensive than those used to treat drug-susceptible TB. Globally, less than half of all patients who start MDR-TB therapy are treated successfully. For the first time in over 40 years, a new TB drug with a novel mechanism of action—bedaquiline—is available, and was granted accelerated approval by the United States Food and Drug Administration in December 2012. There is considerable interest in the potential of this drug to treat MDR-TB. However, information about this new drug remains limited."[22]

Box 2: Evidence to Decision (EtD) framework, question section*

In multidrug-resistant tuberculosis (MDR-TB) patients, should bedaquiline be added to a background regimen based on WHO-recommendations?†

Population: Multidrug-resistant tuberculosis (MDR-TB) patients

Intervention: Bedaquiline plus background MDR-TB treatment

Comparison: Background MDR-TB treatment alone

Main outcomes: Cure by 120 weeks, adverse drug reactions (clinical and biological serious adverse events), mortality, time to culture conversion, culture conversion at 24 weeks, acquired resistance to fluoroquinolone and injectable drugs

Setting: Global, MDR-TB clinics

Perspective: Population perspective (health system)

Subgroups: Patients with extensively drug-resistant (XDR) or pre-XDR tuberculosis or those with resistance or contraindication to fluoroquinolones or injectables

Background:

- The emergence of drug resistance is a major threat to global tuberculosis care and control. WHO estimates that around 310 000 MDR-TB cases (resistant to at least rifampicin and isoniazid) occurred among notified tuberculosis patients in 2011.
- Current treatment regimens for drug-resistant tuberculosis are far from satisfactory: overall duration is ≥20 months, and it requires the daily administration of drugs that are more toxic and less effective than those used to treat drug-susceptible tuberculosis.
- A new drug with a novel mechanism of action—bedaquiline—is available, and was granted accelerated approval by the US Food and Drug Administration in December 2012. However, information about this new drug remains limited

*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.

†Adapted from a WHO guideline.[22] This should not be considered as a WHO recommendation.
An interactive version of this framework which includes subgroup information can be found at http://ietd.epistemonikos.org/#/frameworks/54992ce9352a502d58179c5c/question and at http://dbep.gradepro.org/profile/3879A46D-7E19-FEBA-9B96-BC2B3F996EB1.

Structure of the frameworks

EtD frameworks include three main sections that reflect the main steps in going from evidence to a decision: formulating the question, making an assessment of the evidence, and drawing conclusions. In this article, we illustrate the use of an EtD framework applied to a recommendation about the use of a new drug (bedaquiline) for the treatment of multidrug-resistant tuberculosis (MDR-TB) (box 1, appendix 2).[22] We have used an adapted version of a WHO recommendation as an example.

Formulating the question

The first step in going from evidence to a recommendation or decision is to clearly formulate the question. The question section of an EtD framework includes details of the question in a structured PICO (problem, intervention, comparison, outcomes) format [23]—the perspective from which the options to address the question are considered—relevant subgroups, key background information for understanding the question, and why a recommendation or decision is needed. In the scenario in box 1, the question formulated by the panel was: "In multiple drug-resistant tuberculosis (MDR-TB) patients, should bedaquiline be added to a background regimen based on WHO recommendations?" The panel specified the question details, including the population, intervention, comparison, and outcomes (PICO),[23] and the setting (MDR-TB clinics globally) (box 2). In this example, an adaptation of a WHO recommendation,[22] the panel took a health system perspective, taking into consideration costs (and savings) to the health system and outcomes that might not directly affect the patients being treated.

The perspective that a panel takes will determine which economic consequences of an intervention are considered when making a recommendation or decision. Panels should be explicit about this. It may also affect which outcomes they consider (such as availability and access to health services when considering a health system perspective) and whether they look at equity, acceptability, and feasibility (such as when considering a public health or a health system perspective).

Decisions or recommendations can differ across different subgroups of people. Panels should be explicit about which subgroups they considered, if any, ideally in advance. In the bedaquiline example, the panel paid particular attention to the subgroup of patients with extensive drug resistance and patients with resistance

doi: 10.1136/bmj.i2016 | BMJ 2016;353:i2016 | the bmj

to or contraindications for fluoroquinolones or injectable medications. The rationale was that treatment options for these patients are limited and they may be more likely to accept the risks of a new drug than patients with uncomplicated MDR-TB.

Conflicts of interest

Intellectual and financial conflicts of interest are common and can affect judgments and recommendations or decisions.[24-26] Guideline developers and organisations responsible for healthcare decisions should consider conflicts of interest when a panel is established.[27] In addition, because potential conflicts of interest can vary across questions, panels should consider and report them when formulating each question. They should also specify actions to address these, which can range from simply declaring a conflict of interest to excluding panel members from discussions of specific questions or an entire guideline.[25 27 28] In the bedaquiline example, the panel reported that all panellists declared either minor or no conflicts of interest (appendix 2).

Making an assessment

EtD frameworks make explicit the criteria that are used to assess interventions or options, the judgments made by the panel for each criterion, and the research evidence and additional considerations used to inform each judgment. Research evidence refers to facts (actual or asserted) used to inform the panel's judgments that are derived from studies that used systematic and explicit methods. Additional considerations include other evidence, such as routinely collected data, assumptions, and logic used to make a judgment. Panels may make different judgments for one or more subgroups (such as patients who are older or who have more severe disease) in relation to some or all of the criteria. When relevant, they may also report additional details, such as dissenting views of panel members or the results of voting on judgments where there was disagreement. The assessment of the different criteria made by the panel in the bedaquiline example are available in appendix 2 (an interactive version is available at http://ietd.epistemonikos.org/#/frameworks/54992ce9352a502d58179c5c/question and at http://dbep.gradepro.org/profile/3879A46D-7E19-FEBA-9B96-BC2B3F996EB1.

Different types of decisions and different perspectives require different considerations. Consequently, we suggest specific sets of criteria for clinical recommendations from an individual patient perspective, clinical recommendations from a population perspective, coverage decisions, recommendations and decisions about tests, and health system or public health recommendations and decisions (table 1).

Although there are differences in the operationalisation of the criteria for different types of decisions, most of the criteria are similar, as can be seen in table 1, which shows the criteria for five types of decisions. All five sets of criteria include questions about whether the problem is a priority, the magnitude of the desirable and undesirable effects, the certainty of the evidence, consideration of how patients (or others affected, such as carers) value the main outcomes, the balance between desirable and undesirable effects, resource use, acceptability, and feasibility. All of the frameworks that take a population perspective also include consideration of impacts on equity.

For questions regarding tests, when there is no direct evidence from randomised trials or observational studies of the impact of alternative testing strategies on important outcomes, additional criteria are required.[29] This includes consideration of test accuracy and the certainty of the different types of evidence used to inform judgments about the desirable and undesirable effects of a test (including direct effects, such as adverse effects from invasive tests, and indirect effects, resulting from management decisions based on the test results).

Organisations may want to tailor the criteria that they use. For example, guideline developers may have assessed the priority of problems before making recommendations and therefore might elect not to include the priority of the problem as a criterion. Conversely, some organisations, due to their mandate, might elect to consider a factor separately as an additional criterion rather than as a detailed judgment for a broader criterion. For example, autonomy and other ethical considerations are included as detailed judgments under acceptability in EtD frameworks. However, an organisation might elect to consider autonomy as a separate criterion, rather than as a detailed judgment under acceptability. Table 2 shows other criteria that we have incorporated as detailed judgments, which some organisations might want to consider as separate criteria.

A key feature of the EtD frameworks, like other GRADE-DECIDE presentations,[32] is that they are layered; that is, they present key messages in the top layer with links to more detailed information. For example, the frameworks include concise summaries of the most important research evidence for each criterion (appendix 2). Typically, this is summarised in a table or a paragraph of text. From the framework, it is possible to link to information that is more detailed - for example, an evidence profile [15] or an interactive Summary of Findings table (http://isof.epistemonikos.org/#/finding/543952e4f30d0c47cb1a1495) and from there to even more detailed information, such as a systematic review. This helps to structure discussions, ensure that there is a shared understanding of the key findings of the research that informs each judgment, and avoids problems that sometimes arise when panel members receive large piles of documents without concise summaries. It also makes it easier for panel members and users of recommendations, when needed, to dig deeper into the supporting evidence.

Drawing conclusions

Drawing conclusions begins with the panel reviewing the judgments they have made for all of the criteria in their assessment and considering the implications of those judgments for the recommendation or decision. Based on their assessment, the panel draws conclusions

Table 1 | Criteria for EtD frameworks for five different types of decisions

	Clinical recommendations–individual perspective	Clinical recommendations–population perspective	Coverage decisions	Health system and public health recommendations/decisions	Diagnostic, screening, and other tests*
Priority of the problem	Is the problem a priority?				
Test accuracy	Not applicable				How accurate is the test?
Benefits and harms	How substantial are the desirable anticipated effects?				
	How substantial are the undesirable anticipated effects?				
Certainty of the evidence	What is the overall certainty of the evidence of effects?				What is the certainty of the evidence of: - Test accuracy? - Any critical or important direct benefits, adverse effects, or burden of the test? - Effects of the management that is guided by the test results? - Link between test results and management decisions? - Effects of the test?
Outcome importance	Is there important uncertainty about or variability in how much people value the main outcomes?				Is there important uncertainty about or variability in how much people value the main outcomes, including adverse effects and burden of the test and downstream outcomes of clinical management that is guided by the test results?
Balance	Does the balance between desirable and undesirable effects favour the intervention or the comparison?				Does the balance between desirable and undesirable effects favour the test or the comparison?
Resource use	—	How large are the resource requirements (costs)?			
	—	What is the certainty of the evidence of resource requirements (costs)?			
	Does the cost effectiveness of the intervention (the out-of-pocket cost relative to the net benefits) favour the intervention or the comparison?	Does the cost effectiveness of the intervention favour the intervention or the comparison?		Does the cost effectiveness of the option favour the option or the comparison?	Does the cost effectiveness of the test favour the test or the comparison?
Equity	—	What would be the impact on health equity?			
Acceptability	Is the intervention acceptable to patients, their care givers, and healthcare providers?	Is the intervention acceptable to key stakeholders?		Is the option acceptable to key stakeholders?	Is the test acceptable to key stakeholders?
Feasibility	Is the intervention feasible for patients, their care givers, and healthcare providers?	Is the intervention feasible to implement?		Is the option feasible to implement?	Is the test feasible to implement?

*Tests cover clinical and public health recommendations at individual and population perspectives.

Table 2 | Detailed judgments in Evidence to Decision (EtD) frameworks

Criterion	Detailed judgments
Is the problem a priority?*	• Are the consequences of the problem serious (that is, severe or important in terms of the potential benefits or savings)? • Is the problem urgent? [Not relevant for coverage decisions] • Is it a recognised priority (such as based on a political or policy decision)? [Not relevant when an individual patient perspective is taken]
How substantial are the desirable anticipated effects?	• Judgments for each outcome for which there is a desirable effect
How substantial are the undesirable anticipated effects?	• Judgments for each outcome for which there is an undesirable effect
What is the overall certainty of the evidence of effects?	• See GRADE guidance regarding detailed judgments about the quality of evidence or certainty in estimates of effects[30 31]
Is there important uncertainty about or variability in how much people value the main outcomes?	• Is there important uncertainty about how much people value each of the main outcomes? • Is there important variability in how much people value each of the main outcomes? [Not relevant for coverage decisions]
Do the desirable effects outweigh the undesirable effects?	• Judgments regarding each of the four preceding criteria • To what extent do the following considerations influence the balance between the desirable and undesirable effects: - How much less people value outcomes that are in the future compared to outcomes that occur now (their discount rates)? - People's attitudes towards undesirable effects (how risk averse they are)? - People's attitudes towards desirable effects (how risk seeking they are)?
How large are the resource requirements?†	• How large is the difference in each item of resource use for which fewer resources are required? • How large is the difference in each item of resource use for which more resources are required?
What is the certainty of the evidence of resource requirements?†	• Have all-important items of resource use that may differ between the options being considered been identified? • How certain is the evidence of differences in resource use between the options being considered (see GRADE guidance regarding detailed judgments about the quality of evidence or certainty in estimates)? • How certain is the cost of the items of resource use that differ between the options being considered? • Is there important variability in the cost of the items of resource use that differ between the options being considered?
Are the net benefits worth the incremental cost?*	• Judgments regarding each of the six preceding criteria • Is the cost effectiveness ratio sensitive to one-way sensitivity analyses? • Is the cost effectiveness ratio sensitive to multivariable sensitivity analysis? • Is the economic evaluation on which the cost effectiveness estimate is based reliable? • Is the economic evaluation on which the cost effectiveness estimate is based applicable to the setting(s) of interest?
What would be the impact on health equity?*†	• Are there groups or settings that might be disadvantaged in relation to the problem or interventions (options) that are considered? • Are there plausible reasons for anticipating differences in the relative effectiveness of the intervention (option) for disadvantaged groups or settings? • Are there different baseline conditions across groups or settings that affect the absolute effectiveness of the intervention or the importance of the problem for disadvantaged groups or settings? • Are there important considerations that should be made when implementing the intervention (option) in order to ensure that inequities are reduced, if possible, and that they are not increased?
Is the intervention/option acceptable to key stakeholders?*	• Are there key stakeholders who would not accept the distribution of the benefits, harms and costs? • Are there key stakeholders who would not accept the costs or undesirable effects in the short term for desirable effects (benefits) in the future? • Are there key stakeholders who would not agree with the importance (value) attached to the desirable or undesirable effects (because of how they might be affected personally or because of their perceptions of the relative importance of the effects for others)? • Would the intervention adversely affect people's autonomy? • Are there key stakeholders who would disapprove of the intervention morally, for reasons other than its effects on people's autonomy (such as in regard to ethical principles such as no maleficence, beneficence, or justice)?
Is the intervention feasible to implement?*	For decisions other than coverage decisions: • Is the intervention or option sustainable? • Are there important barriers that are likely to limit the feasibility of implementing the intervention (option) or require consideration when implementing it?[30 31] For coverage decisions: • Is coverage of the intervention sustainable? • Is it feasible to ensure appropriate use for approved indications? • Is inappropriate use (indications that are not approved) an important concern? • Is access to the intervention an important concern? • Are there important legal or bureaucratic or legal constraints that that make it difficult or impossible to cover the intervention?

*The certainty of the evidence could be considered as a detailed judgement for these criteria.
†These criteria are not included when an individual patient perspective is taken.

about the strength of recommendation or type of decision; for example, a strong or weak (sometimes called conditional, discretionary, or qualified) recommendation for or against an intervention or option. In addition, the panel states the recommendation or decision in a concise, clear and actionable manner,[18] and provides the justification for their recommendation or decision. The conclusions also include relevant considerations about subgroups, implementation, monitoring and evaluation, and research priorities (see box 3 for the conclusions reached in the bedaquiline example).

Guideline panels may be reluctant to make a recommendation for or against an intervention or option.

Box 3: Evidence to Decision (EtD) framework—Drawing conclusions section

In multidrug-resistant tuberculosis (MDR-TB) patients, should bedaquiline be added to a background regimen based on WHO-recommendations?*

Recommendation

1. The Expert Group Panel suggests that bedaquiline may be added to a WHO recommended regimen in MDR-TB adult patients under the following conditions [interim conditional recommendation, very low confidence in estimates of effect]:
   - An effective treatment regimen containing four recommended second line drugs from the different classes of drugs according to WHO-recommendations cannot be designed
   - There is documented evidence of resistance to any fluoroquinolone in addition to multiple drug resistance
   - Bedaquiline should be used for a maximum duration of six months and at suggested dosing (400 mg daily for the first 2 weeks, followed by 200 mg three times per week for the remaining 22 weeks)
   - Bedaquiline should be used with caution in people with HIV infection, as well as in patients with comorbidities (such as diabetes) or people with drug or alcohol misuse, because of limited or no information
   - Bedaquiline must not be added alone to a failing regimen.

Justification

The panel formulated a conditional recommendation for a specific subpopulation that can be implemented only under very specific circumstances. The recommendation is also provisional as the panel decided to revise it in 2015 or earlier if substantial data become available increasing the knowledge on safety, toxicity, and efficacy (such as post-marketing studies, ongoing trials and studies). The panel thought that, given the uncertainty about the benefits and harms, only patients with extensive drug resistance tuberculosis would be willing to accept a treatment such as bedaquiline.

The panel judged that the impact on culture conversion was large enough to outweigh the harms for most patients. However, the panel had a low level of confidence in using the available data for global decision making and could not reach consensus on the overall balance of harms and benefits. It proceeded to a vote (observers and technical resources consultants were excluded). The results were as follows: 10 votes that benefits outweighed harms; 4 votes that harms outweighed the benefits; and 2 abstentions (including the chair).

Detailed justification

Undesirable effects—In patients taking bedaquiline, compared with background treatment for MDR-TB, death and serious adverse effects were more common (10 more per 100 patients (from 0 to 89 more) and 5 more per 100 patients (from 0 to 25 more)).

Values—Treatment success, serious adverse events, and mortality were considered important to patients whereas time to conversion, culture conversion, and resistance were less so. Patients with newly diagnosed and proven MDR-TB are unlikely to accept the risk of taking a new drug with potential increase in mortality. The likelihood that patients would accept an effective treatment regimen would depend on the presence of MDR-TB—such as extensively drug-resistant (XDR) or pre-XDR tuberculosis. For this subpopulation there is possibly important uncertainty or variability.

Certainty of the evidence—The overall certainty was low given that the available evidence was limited. There were concerns about imprecision and indirectness due to small sample size, the use of modified intention-to-treat (that is, not intention-to-treat) analysis, and the low quality of evidence for the background MDR-TB treatment regimens used in the trial.

Subgroup considerations

- The likelihood that patients would accept an effective treatment regimen would depend on subgroups of the MDR-TB population: for example, pre-XDR or XDR patient groups may be more likely to accept the risk of taking a new drug with potential increase in mortality than patients with newly diagnosed and proven MDR-TB.

Implementation considerations

- A duly informed decision-making-process by patients should be followed.
- Clinical monitoring and management of comorbidities (especially cardiac and liver disease) should be in place.
- Spontaneous reporting of adverse drug reactions is reinforced at country level, and active pharmacovigilance is established among patient groups treated with the drug.

doi: 10.1136/bmj.i2016 | *BMJ* 2016;353:i2016 | the bmj

(Box 3 continued)

Monitoring and evaluation considerations

- Resistance to other anti-TB drugs should be monitored following WHO recommendations
- In the absence of a specific bedaquiline drug susceptibility test (DST) assay, resistance to bedaquiline should be monitored through assessment of Minimum Inhibitory Concentrations (MICs).
- Baseline testing and monitoring for QT prolongation and development of arrhythmia is imperative.
- Monitor resistance to bedaquiline through assessment of MIC in the absence of a specific bedaquiline DST assay.
- Concerns on scale-up due to costs and/or local regulatory constraints.

Research priorities

- Phase 3 clinical trial(s) of safety and efficacy of bedaquiline, with particular attention to mortality (including causes of death), in the treatment of MDR-TB should be accelerated
- Development of a reliable test for bedaquiline resistance.
- Pharmacokinetics, safety, and efficacy studies in specific populations (children, HIV patients, alcohol and drug misusers, elderly, pregnant women, diabetics, and people with extrapulmonary TB).
- Well designed safety studies events (short and long term), including type, frequency, and severity of adverse events.
- Drug-drug interactions, including with existing and other newly developed anti-TB drugs and antiretroviral drugs.
- Mortality (including cause of death).
- Acquisition of resistance to bedaquiline and to other anti-TB drugs.
- Duration and dosing of treatment.
- Patient acceptability.
- Further research on the validity of culture conversion as a surrogate marker of treatment outcome.

*Adapted from a WHO guideline.[22] This should not be considered as a WHO recommendation. An interactive version of this framework which includes subgroup information can be found at http://ietd.epistemonikos.org/#/frameworks/54992ce9352a502d58179c5c/question and at http://dbep.gradepro.org/profile/3879A46D-7E19-FEBA-9B96-BC2B3F996EB1

Panels should not fail to make a recommendation simply because different people would make different choices. Indeed, that is a defining feature when making a weak recommendation. However, one reason for not recommending for or against an intervention or option is that the pros and cons of the intervention or option and the comparison are so closely balanced that the panel is not prepared to make a weak recommendation in one direction or the other. Another possible reason is that there is so much uncertainty that the panel concludes that a recommendation either for or against the intervention or option would be speculative.[17-19]

The types of recommendations or decisions that are appropriate vary. For example, strong and weak recommendations are appropriate for clinical recommendations and these different types of recommendations have clear implications for clinicians and patients.[17-19] The WHO panel, for example, developed an interim recommendation regarding bedaquiline that was conditional because the certainty of the evidence was very low and because it is recommended only under specific conditions (box 3).

It is not, however, possible to make a strong or weak coverage, health system, or public health decision. For example, an intervention is either covered or it is not, although there can be caveats to coverage. Types of coverage decisions that are possible include not covering an intervention, coverage only in the context of research,[33] covering it with price negotiation, restricted coverage, and full coverage.

The justification for a recommendation or decision should flow from the judgments that the panel made in relation to the criteria used in the assessment. A detailed justification can elaborate on the panel's thinking for the key criteria that drove their recommendation or decision, as illustrated with the bedaquiline example (an adapted version of a WHO recommendation) in box 3. The panel's conclusions about subgroup considerations should specify which subgroups the panel considered and how those considerations affected their recommendation. If the panel's judgments (and the research evidence or additional considerations that informed those judgments) and their conclusions for a subgroup are very different from the overall assessment, the panel can elect to present a separate EtD framework for the subgroup.

Conclusions about implementation considerations should specify key concerns about the feasibility and acceptability of the intervention and strategies to address those concerns, as well as any important information about how to implement the intervention, particularly for complex interventions. Conclusions about monitoring and evaluation should include suggestions for which, if any, indicators should be monitored and any evaluation that is needed in connection with implementing the recommendation or decision. This is particularly relevant for health system and public health decisions and recommendations. Finally, having reviewed and assessed the evidence, panels should identify research priorities to address any important

uncertainties or gaps in the research evidence that informed their judgments.

How are EtD frameworks prepared and used by panels and users of recommendations

Technical teams or others with relevant expertise should generally prepare EtD frameworks. Expertise should typically include an understanding of appropriate systematic review methods,[3 4] the GRADE system,[13 14] and the clinical, health system, or public health topic. The GRADEPro Guideline Development Tool (GRADEPro GDT) (www.gradepro.org), the interactive EtD (http://ietd.epistemonikos.org/), and the interactive Summary of Findings (iSoF; http://isof.epistemonikos.org/) are free, web based software solutions for preparing and using interactive EtD frameworks. The iEtD and iSoF are also integrated in other alternative authoring and publication tools such as MAGIC (Making GRADE the Irresistible Choice; www.magicapp.org). These tools facilitate collaborative preparation and management of EtD frameworks by technical teams and the use of EtD frameworks by panels. They also support the dissemination of information derived from the frameworks to target audiences, including preparation of presentations tailored to clinicians, patients and the public, or policy makers in different formats. GRADEpro also has an all-in-one web solution for managing, summarising, and presenting information for healthcare decision making and developing guidelines. As part of this functionality, GRADEPro GDT supports creating evidence profiles and Summary of Findings (SoF) tables,[15] and it facilitates the development of clinical practice guidelines. GRADEpro also contains a growing database of evidence profiles and evidence to decision frameworks (http://dbep.gradepro.org/search).

EtD frameworks can also be used by guideline developers to adapt recommendations to specific contexts or can be used by decision makers deciding whether to implement a recommendation in their setting. "Recommendation to decision (RtD)" presentations can facilitate this process, as illustrated for the bedaquiline example in appendix 3 (an adapted version of a WHO recommendation). These presentations can be generated by the iEtD. Clinicians and other users of recommendations can use the frameworks to systematically review recommendations and decide whether they are applicable to their setting or to particular patients.

Final remarks and future developments

Over the past 15 years the GRADE Working Group has established criteria for moving from evidence to recommendations. These criteria have been applied in numerous clinical and public health guidelines, and their use has increased transparency in guidelines and provided a structured approach for determining the direction and strength of a recommendation. EtD frameworks are an evolution of this approach to making recommendations.

Advantages of EtD frameworks compared with less structured approaches used in guideline development and decision making include:

- Rigorous development by a wide international multidisciplinary group
- Transparent process for moving from evidence to recommendations or decisions
- Explicit consideration of how much outcomes are valued by those affected by a decision
- Use of a layered approach by panels and in disseminating recommendations or decisions.

The EtD frameworks differ from the earlier versions of GRADE Evidence to Recommendation tables[17-19] in several ways. They incorporate new criteria and require more explicit and structured summaries of evidence to address each criterion, beyond summaries of findings for the effects of interventions.[14] They address coverage, health system, and public health decisions, as well as recommendations, and they facilitate decision making based on recommendations. They require panels to specify the perspective they are taking and differences in their judgments for specific criteria for relevant subgroups. They provide a more detailed structure that can help to facilitate panel discussions, make discussions more efficient, and clarify the research evidence used to inform discussions; and they help ensure that recommendations and decisions flow from judgments about relevant criteria and make the basis for recommendations more transparent.

A potential limitation of EtD frameworks is their increased complexity compared with the previous GRADE Evidence to Recommendation tables. Because healthcare decisions are complex, any system for moving from evidence to decisions requires a balance between simplicity and full transparent consideration of all the important factors. Although EtD frameworks are more complex than the previous approach suggested by the GRADE Working Group for making judgments about the strength of recommendations,[17-19] they add clarity and make the judgments underlying a decision more explicit. Moreover, we have found that, once the question has been formulated and evidence searched for and summarised, the process of reaching decisions using EtD frameworks does not add substantial amounts of time to the decision making process. Nevertheless, as with the use of other methods, mastering the use of EtD frameworks requires familiarisation and practice.

Ideally, research evidence should be used to inform judgments about each criterion in EtD frameworks. However, often research evidence will be lacking or organisations will have limited resources to find and systematically review all of the relevant evidence. EtD frameworks explicitly show what, if any, research evidence was used to inform each judgment and, if no research evidence was available, what considerations were made. Organisations can tailor the criteria that they use and might elect not to use some criteria. However, all of the criteria included in the EtD frameworks can sometimes be critical for a decision. Therefore, we suggest that organisations wanting to reduce the number of criteria, should first consider the implications of doing so. For example, if a guideline developer elects not to include criteria related to resource use, it is then

doi: 10.1136/bmj.i2016 | *BMJ* 2016;353:i2016 | the bmj

either making implicit judgments about resource use or leaving it up to users of their guidelines to consider resource use when deciding whether to adhere to their recommendations.

We have put substantial effort into both identifying a comprehensive set of criteria and making the frameworks as simple as possible. As with all aspects of the GRADE system, we will continue to monitor and evaluate the use of EtD frameworks in practice and, if needed, refine the criteria that are included in each of the frameworks or other aspects of the frameworks.

The use of multiple criteria in making healthcare recommendations or decisions, and the use of evidence that goes beyond evidence of effectiveness and cost effectiveness are not new.[1-4 35-38] Some have argued for the use of multiple criteria decision analysis (MCDA) (using mathematical models) in health technology assessment and coverage decisions.[37-39] However, these models have rarely been used. The advantages and disadvantages of using MCDA compared with EtD frameworks are similar to the advantages and disadvantages of using a balance sheet approach compared with an economic evaluation.[39] It might sometimes be desirable to use both, but few organisations are likely to have the resources to undertake MCDA, and there are many uncertainties regarding MCDA models and their role in informing these types of decisions.

EtD frameworks provide an approach to structured reflection that can help those making recommendations or decisions to be more systematic and explicit about the judgments that they make, the evidence used to inform each of those judgments, additional considerations, and the basis for their recommendations or decisions. For users of recommendations and those affected by decisions, EtD frameworks can help to ensure the trustworthiness of those recommendations or decisions, enable them to appraise the basis for recommendations or decisions, and facilitate adaptation of recommendations or decisions to their own contexts.

AUTHOR AFFILIATIONS

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

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

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

[4]Norwegian Institute of Public Health, Oslo, Norway

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

[6]Department of Internal Medicine, Clinical Epidemiology Unit, American University of Beirut Medical Center, Beirut 1107, 2020, Lebanon

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

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

[9]Department of Medicine, University of Missouri-Kansas City, Kansas City, Missouri, USA, 64110

[10]Evidence Based Health Care Program, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[11]Department of Internal Medicine, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

[12]Epistemonikos Foundation, Santiago, Chile

Members of the GRADE Working Group who contributed to writing this paper include: Pablo Alonso, Holger J Schünemann, Jenny Moberg, Romina Brignardello-Petersen, Elie Akl, Marina Davoli, Shaun Treweek, Reem Mustafa, Gabriel Rada, Sarah Rosenbaum, Angela Morelli, Gordon H Guyatt, Andrew D Oxman, Signes Flottorp, Gerald Gartlehner, Claire Glenton, Joerg Meerpohl, Elena Parmelli, Roger Stanev, Per Vandvik, Ilse Verstijnen, and Erik Von Elm Bjorn. We thank Roman Jaeschke for his comments and suggestions.

We have obtained permission from the WHO to adapt the bedaquiline recommendation used as an example in this article.

Author contributions: PAC and ADO wrote the first draft of this manuscript and PAC, HS, JM, EA, RBP, MD, ST, RM, PV, JM, GHG, ADO, SF, GG, CG, JM, EP, RS, PV, IV, EVE contributed to the writing. All authors have read and commented on the results and conclusions in the manuscript, and have given written agreement of their authorship.

Funding: Work on this article has been partially funded by the European Commission FP7 Program (grant agreement 258583) as part of the DECIDE project. Sole responsibility lies with the authors; the European Commission is not responsible for any use that may be made of the information contained therein.

Competing interests: All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: support from the European Commission FP7 Program (grant agreement 258583); no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work. Authors are members of the GRADE Working Group and the DECIDE project. PAC is supported by a Miguel Servet investigator contract from the Instituto de Salud Carlos III (CPII15/0034).

Transparency: ADO affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

1 Guindo LA, Wagner M, Baltussen R, et al. From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. *Cost Eff Resour Alloc* 2012;10:9. doi:10.1186/1478-7547-10-9.
2 Tromp N, Baltussen R. Mapping of multiple criteria for priority setting of health interventions: an aid for decision makers. *BMC Health Serv Res* 2012;12:454. doi:10.1186/1472-6963-12-454.
3 Fischer KE. A systematic review of coverage decision-making on health technologies-evidence from the real world. *Health Policy* 2012;107:218-30. doi:10.1016/j.healthpol.2012.07.005.
4 Verkerk K, Van Veenendaal H, Severens JL, Hendriks EJ, Burgers JS. Considered judgement in evidence-based guideline development. *Int J Qual Health Care* 2006;18:365-9. doi:10.1093/intqhc/mzl040.
5 Aaserud M, Trommald M, Oxman AD, Innvaer S. [Evaluation of reimbursement applications for new drugs]. *Tidsskr Nor Laegeforen* 2002;122:2619 -23.
6 Harris AH, Hill SR, Chin G, Li JJ, Walkom E. The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004. *Med Decis Making* 2008;28:713-22. doi:10.1177/0272989X08315247.
7 Clement FM, Harris A, Li JJ, Yong K, Lee KM, Manns BJ. Using effectiveness and cost-effectiveness to make drug coverage decisions: a comparison of Britain, Australia, and Canada. *JAMA* 2009;302:1437-43. doi:10.1001/jama.2009.1409.
8 Chambers JD, Morris S, Neumann PJ, Buxton MJ. Factors predicting Medicare national coverage: an empirical analysis. *Med Care* 2012;50:249-56. doi:10.1097/MLR.0b013e318241eb40.
9 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.
10 Alonso-Coello P, Irfan A, Solà I, et al. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. *Qual Saf Health Care* 2010;19:e58.
11 Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet institute of medicine standards: Two more decades of little, if any, progress. *Arch Intern Med* 2012;172:1628-33. doi:10.1001/2013.jamainternmed.56.
12 Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations?*BMJ* 2008;337:a2530. doi:10.1136/bmj.a2530.
13 Atkins D, Best D, Briss PA, et al. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490. doi:10.1136/bmj.328.7454.1490.
14 Guyatt GH, Oxman AD, Vist GE, et al. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-6. doi:10.1136/bmj.39489.470347.AD.

15 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.
16 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.
17 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.
18 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.
19 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.
20 Vogel JP, Oxman AD, Glenton C, et al. Policymakers' and other stakeholders' perceptions of key considerations for health system decisions and the presentation of evidence to inform those considerations: an international survey. *Health Res Policy Syst* 2013;11:19. doi:10.1186/1478-4505-11-19.
21 Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl E, et al. GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: 2. Clinical practice guidelines. *BMJ* 2016;353:i2089. doi:10.1136/bmj.i2089
22 World Health Organization. The use of bedaquiline in the treatment of multidrug-resistant tuberculosis: interim policy guidance. 2013. http://apps.who.int/iris/bitstream/10665/84879/1/9789241505482_eng.pdf.
23 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.
24 Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. *JAMA* 2002;287:612-7. doi:10.1001/jama.287.5.612.
25 Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. *Ann Intern Med* 2010;152:738-41. doi:10.7326/0003-4819-152-11-201006010-00254.
26 Schünemann HJ, Osborne M, Moss J, et al. ATS Ethics and Conflict of Interest Committee and the Documents Development and Implementation Committee. An official American Thoracic Society Policy statement: managing conflict of interest in professional societies. *Am J Respir Crit Care Med* 2009;180:564-80. doi:10.1164/rccm.200901-0126ST.
27 Schünemann HJ, Al-Ansary LA, Forland F, et al. Board of Trustees of the Guidelines International Network. Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. *Ann Intern Med* 2015;163:548-53. doi:10.7326/M14-1885.
28 Boyd EA, Bero LA. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. *Health Res Policy Syst* 2006;4:16. doi:10.1186/1478-4505-4-16.
29 Hsu J, Brożek JL, Terracciano L, et al. Application of GRADE: making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implement Sci* 2011;6:62. doi:10.1186/1748-5908-6-62.
30 Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. GRADE Working Group. What is "quality of evidence" and why is it important to clinicians?*BMJ* 2008;336:995-8. doi:10.1136/bmj.39490.551019.BE.
31 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.
32 Kristiansen A, Brandt L, Alonso-Coello P, et al. Development of a novel, multilayered presentation format for clinical practice guidelines. *Chest* 2015;147:754-63. doi:10.1378/chest.14-1366.
33 Hutton J, Trueman P, Henshall C. Coverage with evidence development: an examination of conceptual and policy issues. *Int J Technol Assess Health Care* 2007;23:425-32. doi:10.1017/S0266462307070651.
34 Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0. Cochrane Collaboration, 2011. http://training.cochrane.org/handbook.
35 Hailey D. Toward transparency in health technology assessment: a checklist for HTA reports. *Int J Technol Assess Health Care* 2003;19:1-7. doi:10.1017/S0266462303000011.
36 EUnetHTA Joint Action 2, Work Package 8. HTA Core Model version 2.0; 2013. Pdf. www.corehta.info/BrowseModel.aspx
37 Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health* 2012;15:1172-81. doi:10.1016/j.jval.2012.06.015.
38 Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Med Decis Making* 2012;32:376-88. doi:10.1177/0272989X11416870.
39 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.

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

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

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;353:i2089
http://dx.doi.org/10.1136/bmj.i2089

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

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.

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

doi: 10.1136/bmj.i2089 | *BMJ* 2016;353:i2089 | the bmj

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

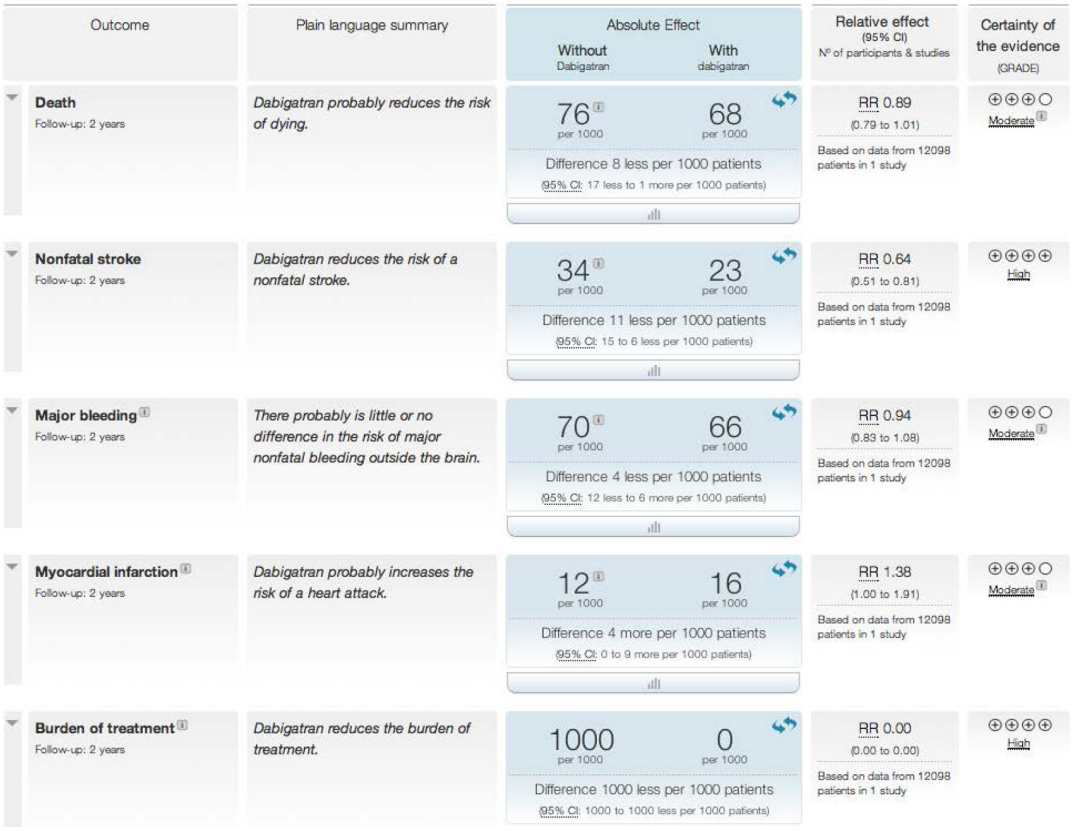


Fig 1 | "Summary of findings" table: dabigatran versus warfarin for atrial fibrillation.[4] An interactive version of this table is at http://isof.epistemonikos.org/#finding/5377108ff30d0c7233205f13

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

doi: 10.1136/bmj.i2089 | *BMJ* 2016;353:i2089 | the bmj

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

Evidence assessment							Summary of findings				
							Study event rates (%)			Anticipated absolute effects	
No of participants (studies) Follow-up period	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Certainty (quality) of the evidence	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]

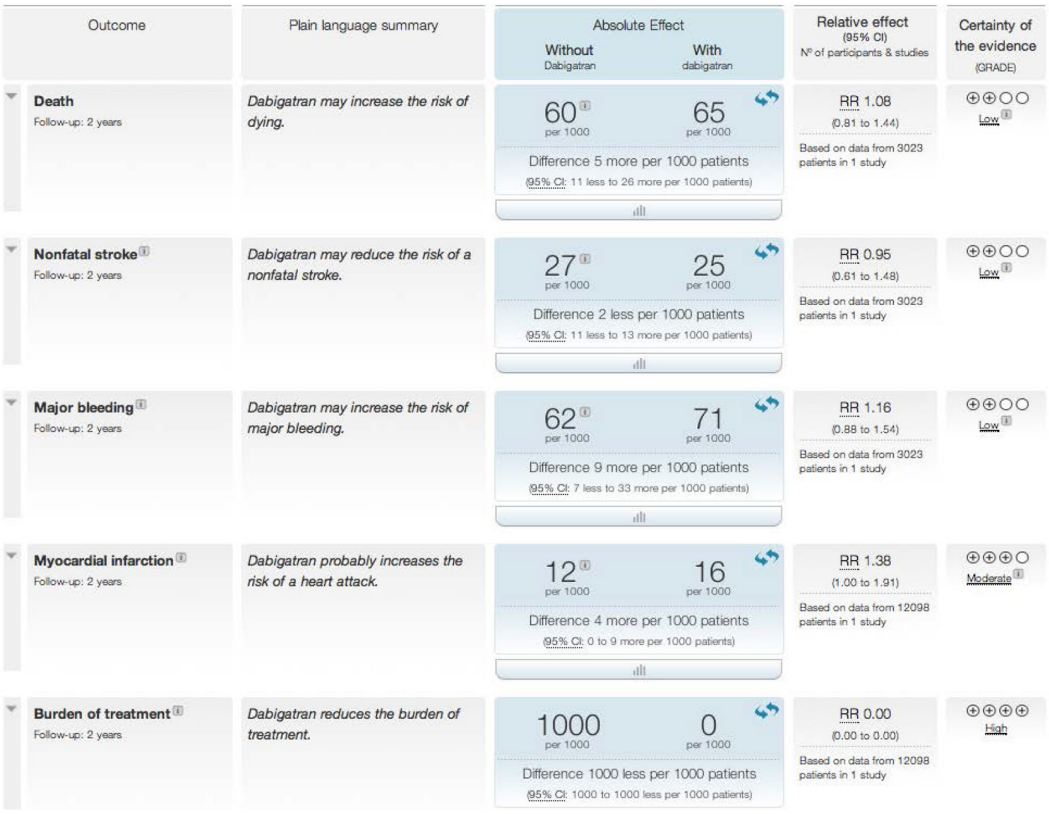
Does the balance between the desirable and undesirable effects favour the intervention or the comparison?

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).



Outcome	Plain language summary	Absolute Effect: Without Dabigatran	Absolute Effect: With dabigatran	Relative effect (95% CI) N° of participants & studies	Certainty of the evidence (GRADE)
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)	⊕⊕◯◯ Low
		Difference 5 more per 1000 patients (95% CI: 11 less to 26 more per 1000 patients)		Based on data from 3023 patients in 1 study	
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)	⊕⊕◯◯ Low
		Difference 2 less per 1000 patients (95% CI: 11 less to 13 more per 1000 patients)		Based on data from 3023 patients in 1 study	
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)	⊕⊕◯◯ Low
		Difference 9 more per 1000 patients (95% CI: 7 less to 33 more per 1000 patients)		Based on data from 3023 patients in 1 study	
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)	⊕⊕⊕◯ Moderate
		Difference 4 more per 1000 patients (95% CI: 0 to 9 more per 1000 patients)		Based on data from 12098 patients in 1 study	
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)	⊕⊕⊕⊕ High
		Difference 1000 less per 1000 patients (95% CI: 1000 to 1000 less per 1000 patients)		Based on data from 12098 patients in 1 study	

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

doi: 10.1136/bmj.i2089 | *BMJ* 2016;353:i2089 | the bmj

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

Members of the GRADE Working Group who contributed to writing this paper include: Pablo Alonso-Coello, Andrew D Oxman, Jenny Moberg, Romina Brignardello-Petersen, Elie Akl, Marina Davoli, Shaun Treweek, Reem Mustafa, Per Vandvik, Joerg Meerpohl, Gordon H Guyatt, Holger J Schünemann, Itziar Etxeandia-Ikobaltzeta, Nora Ibargoyen-Roteta, María Ximena Rojas Reyes, Elena Parmelli, Kevin Pottie, David Rigau, Sarah Rosenbaum, Andrea Juliana Sanabria, Jasvinder Singh, Roger Stanev, Robin W Vernooij and Ilse Verstijnen.

We acknowledge Roman Jaeschke for his comments and suggestions.

Author contributions: PAC and ADO wrote the first draft of this manuscript and PAC, ADO, JM, RBP, EA, MD, ST, RM, PV, JM, GHG, HJS, IEI, NIR, MXRR, EP, KP, DR, SR, AJS, JS, RS, RWV, IV contributed to the writing. All authors have read and commented on the results and conclusions in the manuscript, and have given written agreement of their authorship.

Funding: Work on this article has been partially funded by the European Commission FP7 Program (grant agreement 258583) as part of the DECIDE project. Sole responsibility lies with the authors; the European Commission is not responsible for any use that may be made of the information contained therein.

Competing interests All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: support from the European Comission FP7 Program (grant agreement 258583); no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could seem to have influenced the submitted work. Authors are members of the GRADE Working Group and the DECIDE project. PAC is supported by a Miguel Servet investigator contract from the Instituto de Salud Carlos III (CPII15/0034).

Transparency: PAC affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

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