



# HHS Public Access

Author manuscript

*J Clin Epidemiol.* Author manuscript; available in PMC 2020 June 06.

Published in final edited form as:

*J Clin Epidemiol.* 2018 December ; 104: 103–112. doi:10.1016/j.jclinepi.2018.09.007.

## Evidence to Decision framework provides a structured “roadmap” for making GRADE guidelines recommendations

Shelly-Anne Li<sup>a,\*</sup>, Paul E. Alexander<sup>b</sup>, Tea Reljic<sup>c</sup>, Adam Cuker<sup>d</sup>, Robby Nieuwlaat<sup>b</sup>, Wojtek Wiercioch<sup>b</sup>, Gordon Guyatt<sup>b</sup>, Holger J. Schünemann<sup>b</sup>, Benjamin Djulbegovic<sup>e</sup>

<sup>a</sup>Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, 155 College Street, Toronto, Ontario M5T 1P8, Canada

<sup>b</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, 1280 Main Street W, Hamilton, Ontario L8S 4K1, Canada

<sup>c</sup>Comparative Effectiveness Research Program, Morsani College of Medicine, University of South Florida, 3515 E. Fletcher Avenue, MDT 1212, Tampa, FL 33612, USA

<sup>d</sup>Department of Medicine and Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA

<sup>e</sup>Department of Supportive Care Medicine, Department of Hematology, City of Hope, 1500 East Duarte Road, Duarte, CA 91010, USA

### Abstract

**Objectives:** It is unclear how guidelines panelists discuss and consider factors (criteria) that are formally and not formally included in the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach. To describe the use of decision criteria, we explored how panelists adhered to GRADE criteria and sought to identify any emerging non-GRADE criteria when the panelists used the Evidence to Decision (EtD) framework as part of GRADE application.

**Study Design and Setting:** We used conventional and summative qualitative analyses to identify themes emerging from face-to-face, panel meeting discussions. Forty-eight members from 12 countries participated in the development of five guidelines for the management of venous thromboembolism by the American Society of Hematology.

**Results:** Ten themes corresponded to the GRADE approach and represented all panel discussions. Over half (53%) of the total panel discussions concerned the use of research evidence. When evidence was considered sufficient and clear, the decision-making process proved rapid.

\*Corresponding author. Tel.: +1-416-662-2232; fax: +1-416-978-8222. shellyanne.li@mail.utoronto.ca (S.-A. Li).

Conflicts of interest: The authors report no conflict of interest. S.-A.L. is the lead author of the report, a doctoral candidate. A.C. currently serves on the American Society for Hematology Committee on Quality. R.N., W.W., G.G., and H.J.S. have direct involvement in GRADE methods, working group.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2018.09.007>.

**Conclusion:** The GRADE EtD framework provides structure to guidelines panel meetings, and ensures that the panelists consider all established formal GRADE criteria as they decide on the recommendation text, strength, and direction (for or against an intervention). This is the first study assessing the use of GRADE's EtD framework during real-time guidelines development using panel discussions. Given the widespread use of GRADE, this study provides important information for practice recommendations generated when guidelines panels explicitly follow, in a transparent and systematic manner, the structured GRADE EtD framework. By recognizing the extent to which panels discuss and consider GRADE and other (non-GRADE) criteria for producing guideline recommendations, we are one step closer to understanding the decision-making process in panels that use a structured framework such as the GRADE EtD framework.

### Keywords

Clinical practice guidelines; Decision-making; Epidemiology; GRADE; Evidence to Decision framework; group processing

---

## 1. Introduction

Clinical practice guidelines (CPGs) are systematically developed statements to assist clinicians in making health care decisions that optimize patient care and a valuable source of guidance when based on best available evidence [1]. Clinicians need guidance from trustworthy recommendations that are underpinned by transparent ratings of certainty of evidence and other relevant factors [2]. Grading of Recommendations Assessment, Development, and Evaluation (GRADE) has emerged as the leading system for rating the certainty of evidence and strength of recommendations in guidelines development [2–5]. The GRADE approach provides explicitness, transparency, and structure for judgments regarding the certainty of research evidence—and other factors—such as the magnitude of benefits and harms, patient values and preferences, resource and cost considerations, acceptability, feasibility, and health equity to offer guidance on grading the strength of recommendations as strong or conditional (for or against intervention of interest) [2–7].

### 1.1. Evidence to Decision frameworks

Evidence to Decision (EtD) frameworks are developed to assist panels in using evidence in a more structured and transparent manner to inform decisions in the context of clinical recommendations (treatment and diagnosis), and health system or public health recommendations [8]. In using EtD frameworks as part of GRADE application, the aim is to ensure that all important criteria are considered in the context of guideline/recommendation development. EtD frameworks include three sections that represent the key steps to evidence-informed decision-making: formulating the question, assessing the evidence and other relevant factors identified by GRADE, and drawing conclusions (Appendix 1) [8].

### 1.2. Group decision-making in clinical practice guidelines panels

CPGs are a product of decision-making by a group of 10–20 individuals with varying expertise in content, research methodology, clinical practice, and patient experience (e.g., patient representatives). Although judgment criteria and decision-making processes during

guidelines development serve as the cornerstone for trustworthy guidelines [1], group decision-making processes of guidelines panelists remain poorly documented [9–14].

Theories of decision-making are broadly classified as normative (describes how people “should” or “ought to” make decisions) and descriptive (identifies how people actually make their decisions) [15]. EtD frameworks provide normative instructions that support decisions about how CPG panels “should” or “ought to” develop clinical practice recommendations [16].

However, normative models have been repeatedly shown to be inconsistent with the actual decision-making process [15–19]. Many factors are known to affect how people make their decisions. These factors can be broadly categorized as those related to a) decision features or characteristics of the decision/recommendation itself (e.g., consideration of equity when developing guidelines for vulnerable populations in a politically charged atmosphere) [20]; b) situational/contextual factors (e.g., clinical setting, time pressure, cognitive load, social context, gender bias, defensive medicine); c) individual characteristics of the decision-maker (e.g., role on the panel, cultural background, race/ethnicity, professional background, methodological expertise) [21,22]; and d) the role of emotions (e.g., regret regarding potentially wrong recommendations) [23,24]. It is unclear if and to what extent these or any other factors contribute to the decision-making process in the development of guidelines.

### 1.3. Study aim

We aimed to determine whether GRADE (normative criteria) dominate non-GRADE (descriptive factors). While some empirical studies quantitatively assessed the association between GRADE factors and strength of recommendations in guidelines development meetings [25,26], few described how GRADE and non-GRADE factors are considered by panelists in such meetings. To our knowledge, no study has analyzed the decision-making process of guidelines panelists who applied an EtD framework for making GRADE guidelines recommendations using real-time deliberations from guidelines development meetings. In this study, we aimed to understand the ways in which panelists used the GRADE EtD framework, and the aspects of the framework in which they focused, by analyzing the deliberations among the panelists during the guidelines development panel meetings. This is the first report from among a larger study that aims to describe and explain how decision-making factors of the panelists may influence clinical practice recommendations.

## 2. Methods

We used a qualitative descriptive approach to classify the themes that panelists discussed during recommendations development. The analysis can be viewed as purely descriptive with no causality inferred. As a background, each of the included panel meetings spanned 2 days and had a preset agenda that guided the panelists through each day. A guideline panel chair and co-chair led the panels that included content experts, methods experts, epidemiologists, clinicians, researchers, and patient representatives. Before the meetings, panelists underwent intensive training on using the EtD framework when making GRADE guidelines recommendations.

## 2.1. Data collection

We audio-recorded deliberations of five panels convened by the American Society Hematology (ASH) to develop guidelines for the management of venous thromboembolism (VTE). Deliberations were transcribed verbatim by a professional contractor. All participants' names, or any names and affiliations mentioned during the panel discussion, were de-identified in the transcripts to ensure anonymity.

## 2.2. Data analysis

We used both conventional and summative approaches to qualitative analysis [27]. For conventional analysis, we used inductive and deductive content analyses. The deductive analysis allowed us to determine a priori themes based on the GRADE EtD framework, which includes key criteria (certainty of evidence, cost/resource use, balance of benefits and harms, patient values and preferences, acceptability, feasibility, and health equity) that panelists ought to take into consideration when producing clinical practice recommendations. We used inductive content analysis to uncover other non-GRADE themes that emerged from the data [27].

The conventional analysis proceeded in several steps. First, to get a sense of the discussions, two investigators (S.-A.L., P.E.A.) read the transcripts of all panel meetings. Second, deductive content analysis began with reading and then highlighting all segments of coherent text units (i.e., at least one sentence) that could be mapped onto the GRADE EtD framework. Third, these text units were assigned with a condensed label (code) that captured the essence of the unit's meaning. Fourth, we coded all the highlighted passages and categorized these codes into the corresponding a priori themes. Finally, we searched for any new information not already captured by using inductive content analysis. Any content that did not fit into a priori themes was assigned a new code. Both investigators (S.-A.L., P.E.A.) conducted the steps independently and in duplicate. Reviewers resolved discrepancies by discussion.

We used summative content analysis to explore the usage of a word or a group of words by analyzing the frequency and content of the deliberations. Summative content analysis was performed using NVivo 11 [28] to execute the search strategy (Appendix B). Before reading the transcripts, we developed a search strategy with an academic librarian. We amended additional terms to the search strategy after reading through the transcripts. For example, "Africa" was added as a search term in the theme Accessibility because panelists considered how the treatment could be accessed by individuals in Africa. Introductory discussions on the panelists' roles and names, dialog about technical difficulties that occurred during panel meetings about accessing document files, and concluding remarks were omitted from the analysis.

We summarized the number of occurrences from each key term or phrase and its alternative terms and phrases. In the summative approach, counting identifies patterns in the data and helps contextualize codes [29], and allows for the interpretation of context associated with the use of the word or phrase [27]. We computed the percentage of words devoted to each

theme. Two investigators (S.A.L., B.D.) conducted the summative analysis independently and in duplicate, and also discussed interpretations with a third investigator (P.E.A.).

Our objective was to describe how panelists consider topics that clearly match on GRADE criteria and those that do not during group decision-making processes. In this report, we focused on reporting the results of the summative content analysis. The University of South Florida IRB (IRB#: Pro00027571) approved the project.

### 3. Results

The decision-making panels consisted of 48 members (40 content experts and methodologists, 8 patient representatives) from Belgium ( $n = 2$ ), Canada ( $n = 12$ ), the United States ( $n = 21$ ), Germany ( $n = 2$ ), Italy ( $n = 1$ ), the United Kingdom ( $n = 1$ ), Brazil ( $n = 1$ ), Austria ( $n = 2$ ), Australia ( $n = 2$ ), Denmark ( $n = 1$ ), the Netherlands ( $n = 3$ ), and Switzerland ( $n = 1$ ). The content experts and evidencebased methodologists/epidemiologists represented nine sub-specialty areas (anesthesia, hematology, internal medicine, obstetrics, oncology, cardiology, surgery, pharmacy, and emergency medicine). The panelists participated in meetings to make judgments and reach consensus decisions for CPGs on each of the following panels: (a) prevention of VTE in surgical hospitalized patients, (b) prevention of VTE in medical patients, (c) prevention of heparin-induced thrombocytopenia (HIT), (d) prevention of thrombophilia, and (e) prevention of VTE in the context of pregnancy. Before these meetings, all guidelines panels had regular panel calls to prioritize guideline questions and outcomes, discuss approaches for synthesizing evidence, review evidence summaries and systematic review progress, and in some cases, to discuss prevoting results for guideline questions.

The mean age of panelists was 49.9 years ( $SD = 10.2$ ), and mean years of professional experience for panel experts was 19.1 years ( $SD = 6.3$ ). There were 23 (47.9%) females. Most experts ( $n = 34$ ; 57.6%) rated their level of expertise as higher than other experts on the panels, 10 (20.4%) rated their level of expertise as the same as others, and the remaining ( $n = 5$ ; 10.2%) as lower than others. The panel meetings were held in November 2016 in Washington, D.C., United States. All panel meetings were conducted in English.

Panel meetings lasted between 15 and 26 hours across a span of 2 days (median = 11 hours per day), for a total of 100 meeting hours. Verbatim transcripts ranged from 82 to 348 pages per day, totaling 2469 pages of transcription. A total of 10 themes (Fig. 1) emerged. Table 1 presents the themes and their corresponding supporting quotes/phrases.

#### 3.1. GRADE themes

We define a “GRADE theme” as any theme that was listed as a GRADE EtD framework criterion or as an aspect to consider (as per instructions) when using the formal GRADE approach to produce recommendations. Of the 10 GRADE themes, six directly mapped onto the GRADE EtD framework criteria: a) research evidence, b) resource use and costs, c) balance of benefits (desirable outcomes) and harms (undesirable outcomes), d) patient values and preferences, e) health equity, and f) feasibility and acceptability (as one theme). Other GRADE EtD criteria such as priority of the problem did not emerge as a theme with

deeper or extended discussions because this criterion was reportedly addressed in panel calls before in-person panel meetings.

Research evidence, which emerged as a prominent theme routinely discussed in panels, was both a GRADE EtD framework criterion (certainty of evidence) and a criterion for making judgments for all other EtD framework criteria. For example, evidence was used to inform decisions about the balance between desirable and undesirable effects for an intervention or treatment. However, most discussion related to research evidence was about the clinical effects of an intervention or treatment.

In addition, four themes (clinical experience, legal implications, political environment, and language in guideline) emerged from the inductive content analysis, with a combined contribution of 6% of the total panel discussions. Although these themes are acknowledged by GRADE [1–7] (i.e., clinical experience constitute a necessary source of evidence when published evidence is unavailable), they are not explicitly incorporated as criteria within the GRADE system. We present the findings for all themes in detail in the following section.

### 3.1.1. Research evidence on the clinical effects of a treatment or intervention

Discussion of research evidence to determine clinical effects of a treatment or intervention was prominent across all panels (occupied 53.1% of all discussions). When research evidence was considered sufficient and clear, the decision-making process was rapid; the overall deliberation totaled only of 3.7% of discussions. The lack of evidence or low-quality evidence was a recurrent issue and occupied 12.8% of the discussion topics. For example, panelists recognized the limitations of their recommendations due to low-quality evidence:

I do see that you make an enormous amount of work, knowing the literature, how more or less, we will probably end up with very small numbers, very weak evidence, not being able to make any firm conclusion out of those data.

(HIT Panel)

Sources of uncertainty, imprecision (denoted by wide confidence intervals and small sample size and number of events), occupied 2% of the discussions, whereas 5.1% focused on indirectness of data (e.g., surrogate outcomes, lack of similarity between study population and guideline target population, and outcomes). The remaining discussions (29.4%) about evidence were found in verbal summaries of study findings to guide the panel discussions.

**3.1.2. Resource use and costs**—Resource use and costs occupied 13.4% of all discussion topics. The panelists considered the affordability of the treatment or drug in multiple countries at the levels of the patient and the local government. When deciding on recommendations, the panelists reflected on how local governments in various nations may interpret guidelines and make decisions based on the existing economic status of their nation.

They also considered the cost-effectiveness of the drug or treatment—that is, not only the cost of the drug or therapy itself but also the subsequent costs of treatment downstream (e.g., prolonged stay in the intensive care unit [ICU]):

About 80% of the costs are due to prolonged in-hospital stay in ICU treatment. The drug costs by comparison are relatively small. So any drug that will reduce intensive care treatment of outpatients due to more problematic complications will always be cost-effective regardless of what the price is.

(VTE in Surgical Patients Panel)

Panelists also considered relevant cost-effectiveness analyses identified in the literature, and suggested making cost-effectiveness models that could better inform their recommendations:

...we know how much it costs, testing and treating. But what is the blended result if we start pretesting, treating and we take into account also [the] visit to the patient and saved events or caused events, and we make a formal cost-effectiveness model.

(Thrombophilia Panel)

**3.1.3. Balance of benefits and harms**—Balance of benefits and harms was a routine component of panel deliberations, occupying 12.5% of the deliberations. The panelists provided different scenarios and then discussed the impact on reductions in mortality and the benefits on other outcomes, and considered adverse events. One content expert in the thrombophilia panel suggested, "...okay let's assume every thrombophilia in every clot. What's the impact on mortality...what was the impact on whatever other outcomes we have down the road." Similarly, a content expert on the pregnancy panel grasped the trade-off between benefits and risks, offering points for consideration to other panelists when discussing about balance of benefits and risks/harms:

And based on what we know, even if this study shows a moderate benefit, there are still risks to the patient population that we are dealing with. That would mean that you couldn't just transcribe that potential benefit.

(Pregnancy Panel)

**3.1.4. Feasibility and acceptability**—This theme occupied 11.4% of the discussion topics. For instance, the panelists discussed practicality, reflecting on the logistics and how recommendations for specific conditions would be integrated into health organizations and systems:

These guidelines are changing so we came from the situation of wanting to recommend every patient on heparin, but with the increasing early dismissal from hospital, it becomes impractical doing this.

(HIT Panel)

Acceptability was also an important decision factor. For instance:

Yes so I'm not going to make a recommendation because I am conflicted but I think in this case the acceptability of the treatment is actually pretty important because in this time basically everybody takes an anticoagulant or drug after they go home.

(VTE in Surgical Patients Panel)



**3.1.5. Patient values and preferences**—This theme occupied 3% of all the discussion topics. Panel chairs typically consulted patient representatives as a source for assessment of patient values and preferences when patient values were the topics of interest. To illustrate, a patient representative from one of the panels was asked for their perspective on doing an ultrasound to assess for silent deep vein thrombosis (DVT) and extending anticoagulation for 3 months if silent DVT is identified. The patient representative vocalized the benefits and harms of the procedures:

From my point of view, I don't see why you wouldn't do it. If there is a 50% or 40% chance that I could have a DVT, then I want to be anticoagulated because I want to know about it. And an ultrasound is [a] noninvasive thing, it is not a dangerous procedure; there is no down-side from it as far as I can see. What's the harm in doing that if there is a lot of benefit?

(HIT Panel)

Content experts also shared their clinical expertise and experiences on patient values. The experts considered the extent to which recommendations were consistent with the values of the patients. For instance, one expert shared their thoughts on pregnant women's desire to receive low-molecular-weight heparin,

Now that [statement in the recommendation] is fine because at the moment there is clearly a desire on the part of women, and treaters to give low molecular weight heparin; and what we are not doing is we are not actually depriving any woman of getting that by this statement.

(Context of Pregnancy Panel)

**3.1.6. Health equity**—This theme occupied 1.5% of the deliberations. The theme/term “health inequity”, or an uneven distribution of access to tests, medications, treatments, or interventions, was considered across all panels. The panelists deliberated differences in treatment access due to variations in medical coverage regulations. For instance, an HIT panel member commented “So those who don't have coverage won't have access to that very good drug and we could be creating some inequity or greater inequities in certain countries.” Another panelist focused on treatment availability and reflected:

The US healthcare system, if you're going to get into that, say your treatment and it cost you \$2000 and you don't have coverage for, that's going to increase health inequity.

(Thrombophilia Panel)

**3.1.7. Language in guideline**—Language in guideline represented approximately 3% of the discussions across all panel meetings. The panelists were inclined to be transparent about how they made decisions and to communicate any assumptions, concerns, reservations, or conflicting viewpoints that were raised during the decision-making process. To illustrate, one panel member suggested the wording of how recommendations were made:

We should add, ‘For one recommendation based on the data and clinical experience that were available, the panel was unable to make a recommendation.’ And we



might actually add some language about that; there were some panelists who favoured choice one and others who favoured choice two?

(HIT Panel)

**3.1.8. Legal context**—Legal implications represented 1% of panel discussions. The panelists considered how certain recommendations may put physicians at risk of future lawsuit. They also reflected on how lawyers might interpret the guidelines when making their legal cases, and considered the importance of making recommendations that would protect physicians from legal implications while ensuring patient safety. The panelists were also cognizant that the ASH guidelines may be adopted as national policies. For example, “I think your point that the cost of treating 12 months vs. a strategy of testing and then treating selectively might make some jurisdictions decide, you know, well that’s going to be our policy because that’s all our health care system can afford.” (Thrombophilia Panel).

**3.1.9. Clinical experience**—Clinical experience/gestalt represented 1% of the total panel discussions. The panelists agreed that clinical experience or gestalt alone would not be a sound judgment when making recommendations. They suggested that clinical experience should always be accompanied by research evidence. One panelist suggested, “So...I think that we have come around to the idea together that we do recommend a 4T score [pretest scoring system to identify HIT patients] in addition to or after gestalt rather than gestalt alone.”

The panelists brought their clinical experiences to brainstorm with other members about making the best recommendation. One panel member shared,

So, I think another important point here is just again from clinical experience that when you talk about using warfarin for postpartum prophylaxis for 6 weeks with patients, the hassle of getting INR [international normalized ratio] testing with a newborn at home and the fact that you’re only going to be on it for 6 weeks, you will get to a stable dose at the end of the treatment.

(Pregnancy Panel)

**3.1.10. Political environment**—Considerations of the political environment represented 1% of the total panel discussions. The panelists were aware of the political implications when making specific recommendations and drew on relevant political advances to help inform guidelines decision-making. For instance,

The lateral example in Australia—the risk of getting HIV from a blood transfusion is approximately 1 in 20 million, but such was the community uproar that the government invested \$20 million a year to introduce additional testing to make that risk lower.

(Non-Surgery Panel)

## 4. Discussion

When used in a structured, explicit, transparent, and well-defined format such as the EtD framework, panelists faithfully adhere to GRADE. Explicit GRADE criteria occupied 94% of the panel discussions, whereas four themes that were not part of the explicit GRADE criteria (clinical experience, legal implications, political environment, and language in guideline) contributed to 6% of the total panel discussions. Use of research evidence to inform decisions was apparent across all GRADE EtD criteria, contributing to >50% of the discussions. The panelists used research evidence substantially more frequently to inform recommendation decisions than clinical experience (<2% of the discussion topics). When making judgments for recommendations in the presence of clear evidence, panelists pay far more attention to evidence than on personal clinical experiences.

Although the discussion of GRADE themes dominated the CPG meetings, findings do not establish that descriptive (experiential/clinical expertise and other factors) as opposed to normative factors play no role in the guidelines development process. Rather, the findings are more likely evidence of the instruction effect; when people are given specific instructions to deliberate within given normative rules, descriptive responses can be suppressed [15]. This postulation is highly relevant to this study because these panelists underwent extensive training on using the EtD framework as part of GRADE application. When panelists are trained and coached to follow certain instructions according to specific rules, they tend to follow such rules. Sundberg et al. [30] found that panelists, with clear instructions for the use of research evidence for guidelines development, would adhere to these instructions. In our study, considerations of the language used in the guideline, clinical experience, legal context, and the national political environments emerged but were not part of the formal GRADE criteria. It is possible that, without the extensive preparatory training and the EtD framework, these themes may be more prominent and expanded on in the discussions. In our study, the development of guidelines occurred within explicitly identified GRADE EtD criteria and other factors that are acknowledged as important within the framework [8].

### 4.1. Strengths and limitations

Investigators analyzed the transcripts independently and in duplicate, thus mitigating bias. We used a summative approach for qualitative content analysis, requiring the least amount of subjective interpretation from researchers [27]. This approach generates basic insights into what, and how, words or groups of words related to GRADE and non-GRADE themes are used during the decision-making process of developing guidelines recommendations when applying the GRADE EtD framework.

Our analysis was limited to the panel meeting discussions; we did not evaluate the premeeting conference calls, meetings, and training material that provided information about how to use the EtD framework. We also had no access to the extent of EtD training and whether that training was uniform across panel groups. This information may explain, to some extent, why certain topics had less discussion. Reasons explaining why there is emphasis on certain framework criteria, and the identification and prioritization of questions, remain unknown. Thus, our study relates to only the discussion that had occurred during guidelines panel meetings.

On a final note, the summative approach used in this analysis precludes a richer understanding of the data. For instance, it is unclear how GRADE criteria such as acceptability, resource use and costs, and patient values are considered in light of the quality of evidence and the local context to which the recommendations will be translated.

#### 4.2. Relation to prior work

Investigators have made efforts to understand the experiences of methodologists and panelists who applied EtD frameworks when developing clinical practice recommendations. For example, Neumann et al. [31] solicited feedback from methodologists and panelists about the use of EtD frameworks, and Dahm et al. [32] sought narrative feedback from international health policymakers and other stakeholders on the applicability of the GRADE EtD framework for coverage decisions. However, narrative feedback and interviews are influenced by social structural characteristics (e.g., social context, socioeconomic positioning of informants, ethnicity, gender) and the framing of the questions. These characteristics may inhibit information disclosure and contribute to the construction of an account that lends itself to particular responses [28]. Our analysis of real-time, face-to-face, guidelines development meetings mitigated these limitations.

In keeping with the findings from Neumann et al. [31], we found that the GRADE EtD framework provides a highly structured and explicit way to guide panel decisions on recommendations. The finding that panelists place emphasis on evidence when making recommendations rather than using clinical experience is congruent with a recent qualitative inductive longitudinal case study approach investigating the decision-making process of CPGs for disease prevention [30].

Health equity and patient values were the least discussed of all GRADE themes. This is congruent with existing evidence, which suggests that the consideration of patient values is inconsistent in panel deliberations and guidelines reporting and, in many cases, absent [33–35]. This does not mean that the panels thought that the values and preferences were not important; rather, the extent of the discussions was dominated by other factors. Past research suggests that patient values are ultimately constructed during the physician-patient encounter [34], which becomes difficult to assess in a noncontextual environment (e.g., guidelines panel). Preparations before the panel meetings that were related to patient values and preferences may partially explain why they were not substantially discussed in the panel meetings.

#### 4.3. Implications

Although GRADE may comprehensively cover all important issues relevant to guidelines decision-making, the EtD structure may have an important influence on the time spent on different criteria. For instance, were legal issues included as a seventh explicit criterion within the GRADE system rather than simply as an issue to be considered within the acceptability and feasibility criteria, panels may spend more time on the issue than they did. Whether that would represent good use of the panels' limited time remains uncertain, however. Both rigor and efficiency are important aspects of the guidelines development, and the EtD framework seems to provide a structured, concise, and transparent approach to help

bridge evidence to health care decisions, including recommendations by guidelines panels. Future research should examine how GRADE guidelines panels that do not have extensive training and access to the EtD framework would consider non-GRADE criteria when making recommendations.

For those who believe that GRADE comprehensively captures all the factors that panels should be considering in making recommendations, the results are in general reassuring—the EtD framework promotes structure and adherence to the GRADE approach when issuing recommendations. Alternatively, for those who believe that key elements are missing from the GRADE approach, following the EtD framework may suppress consideration and discussion of those missing elements. For example, Morgan et al.'s critical interpretive synthesis of 19 published frameworks to guide decision-making for guidelines developers recommended that GRADE EtD framework developers officially broaden the acceptability and feasibility criteria to explicitly include political and health system factors, which would increase the applicability in diverse political and health systems contexts [36].

## 5. Conclusions

This is the first study assessing the use of the GRADE EtD framework via real-time, recorded panel deliberations. Given the widespread use of GRADE, this study provides important information for practice recommendations generated when guidelines panels explicitly follow, in a point-by-point manner, the structured GRADE EtD framework. Group decision-making during guidelines development is a dynamic and complex process [36]. By recognizing the extent to which panels discuss and consider GRADE and other (non-GRADE) criteria for producing guidelines recommendations, we are one step closer to understanding the decision-making process of panelists who use a structured framework such as the GRADE EtD framework. Future research can investigate other factors that contribute to guidelines decision-making. Existing literature suggests that group composition and communication styles, contextual/situational factors, and individual characteristics of group members all contribute to group decisions, and thus, warrant further exploration [36–39]. To investigate how panel decision-making unfolds when considering these factors, we used conventional content analyses to gain deeper insights through both explicit and inferred communications that occur between panelists, as well as mixed-effect quantitative modeling. These findings will be reported in subsequent research articles.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors thank Robert Kunkle and his staff from the American Society of Hematology (ASH) for helping facilitate our project. The authors also wish to thank all the panelists participating in development of the ASH guidelines on venous thromboembolism for agreeing to take part in our study. They also thank the ASH leadership for supporting this project. ASH supported the overall project of guideline development for the ASH VTE guidelines including the development of the EtD frameworks to Dr Schünemann through funding for McMaster University.

This project was supported by grant number R01HS024917 from the Agency for Healthcare Research and Quality for PI: Dr Djulbegovic. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. Funding sources have no involvement in any of the activities conducted in this article.

## References

- [1]. Eccles MP, Grimshaw JM, Shekelle P, Schünemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implementation Sci* 2012;7:60.
- [2]. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al., GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6. [PubMed: 18436948]
- [3]. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6. [PubMed: 21208779]
- [4]. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, 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. [PubMed: 23570745]
- [5]. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94. [PubMed: 21195583]
- [6]. GRADE working group. Available at <http://www.gradeworkinggroup.org/>. Accessed June 22, 2017.
- [7]. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, 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. [PubMed: 22542023]
- [8]. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE evidence to decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ* 2016;353:i2089. [PubMed: 27365494]
- [9]. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines international network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;156:525–31. [PubMed: 22473437]
- [10]. Garcia-Alamino JM, Ward AM, Alonso-Coello P, Perera R, Bankhead C, Fitzmaurice D, et al. Self-monitoring and self-management of oral anticoagulation. *Sao Paulo Med J* 2010;128(4):246. [PubMed: 21120440]
- [11]. Oxman AD, Fretheim A, Schünemann HJ. Improving the use of research evidence in guideline development: introduction. *Health Res Pol Syst* 2006;4(1):12.
- [12]. Kunz R, Fretheim A, Cluzeau F, Wilt TJ, Qaseem A, Lelgemann M, et al. Guideline group composition and group processes: article 3 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proc Am Thorac Soc* 2012;9(5):229–33. [PubMed: 23256164]
- [13]. Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Res Policy Syst* 2006;4:15. [PubMed: 17134482]
- [14]. Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 5. Group processes. *Health Res Policy Syst* 2006;4:17. [PubMed: 17140442]
- [15]. Bell DE, Raiffa H, Tversky A, editors *Decision making: descriptive, normative, and prescriptive interactions*. Cambridge, UK: Cambridge University Press; 1988.
- [16]. Djulbegovic B, Elqayam S. Many faces of rationality: implications of the great rationality debate for clinical decision-making. *J Eval Clin Pract* 2017;23:915–22. [PubMed: 28730671]
- [17]. Djulbegovic B, Elqayam S, Dale W. Rational decision-making in medicine: implications for overuse and underuse. *J Eval Clin Pract* 2018;24(3):655–65. [PubMed: 29194876]

- [18]. Logic Evans J. and human reasoning: an assessment of the deduction paradigm. *Psychol Bull* 2002;128(6):978–96. [PubMed: 12405140]
- [19]. Heit E, Rotello CM. Traditional difference-score analyses of reasoning are flawed. *Cognition* 2014;131(1):75–91. [PubMed: 24462712]
- [20]. US Preventive Service Task Force. Screening for breast cancer: U.S. preventive services task force recommendation statement. *Ann Intern Med* 2009;151:716–26. [PubMed: 19920272]
- [21]. Appelt KC, Milch KF, Handgraaf MJJ, Weber EU. The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment Decis Making* 2011;6:252–62.
- [22]. Djulbegovic B, Beckstead JW, Elqayam S, Reljic T, Hozo I, Kumar A, et al. Evaluation of physicians' cognitive styles. *Med Decis Making* 2014;34:627–37. [PubMed: 24722474]
- [23]. Djulbegovic B, Frohlich A, Bennett CL. Acting on imperfect evidence: how much regret are we ready to accept? *J Clin Oncol* 2005;23:6822–5. [PubMed: 16145058]
- [24]. Djulbegovic M, Beckstead J, Elqayam S, Reljic T, Kumar A, Paidas C, et al. Thinking styles and regret in physicians. *PLoS One* 2015;10:e0134038. [PubMed: 26241650]
- [25]. Kumar A, Milandinovic B, Guyatt GH, Schunemann HJ, Djulbegovic B. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines. *J Clin Epidemiol* 2016;75:115–8. [PubMed: 26845745]
- [26]. Djulbegovic B, Kumar A, Kaufman RM, Tobian A, Guyatt GH. Quality of evidence is a key determinant for making a strong GRADE guidelines recommendation. *J Clin Epidemiol* 2015;68:727–32. [PubMed: 25766057]
- [27]. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15:1277–88. [PubMed: 16204405]
- [28]. NVivo qualitative data analysis Software. Melbourne, Australia: QSR International Pty Ltd. Version 11; 2015.
- [29]. Morgan DL. Qualitative content analysis: a guide to paths not taken. *Qual Health Res* 1993;3:112–21. [PubMed: 8457790]
- [30]. Sundberg LR, Garvare R, Nyström ME. Reaching beyond the review of research evidence: a qualitative study of decision making during the development of clinical practice guidelines for disease prevention in healthcare. *BMC Health Serv Res* 2017;17:344. [PubMed: 28490325]
- [31]. Neumann I, Brignardello-Petersen R, Wiercioch W, CarrascoLabra A, Cuello C, Akl E, et al. The GRADE evidence-to-decision framework: a report of its testing and application in 15 international guideline panels. *Implement Sci* 2016;11:93. [PubMed: 27417219]
- [32]. Dahm P, Oxman AD, Djulbegovic B, Guyatt GH, Murad MH, Amato L, et al. Stakeholders apply the GRADE evidence-to-decision framework to facilitate coverage decisions. *J Clin Epidemiol* 2017;86:129v39. [PubMed: 28377194]
- [33]. McCormack JP, Loewen P. Adding “value” to clinical practice guidelines. *Can Fam Phys* 2007;53(8):1326–7.
- [34]. Burgers JS, Cluzeau FA, Hanna SE, Hunt C, Grol R. Characteristics of high-quality guidelines. *Int J Technol Assess Health Care* 2003; 19(1):148–57. [PubMed: 12701947]
- [35]. Zhang Y, Coello PA, Broek J, Wiercioch W, Etxeandia-Ikobaltzeta I, Akl EA, et al. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. *Health Qual Life Outcomes* 2017; 15(1):52. [PubMed: 28460638]
- [36]. Aritz J, Walker RC. Cognitive organization and identity maintenance in multicultural teams: a discourse analysis of decision-making meetings. *J Business Commun* (1973) 2010;47(1):20–41.
- [37]. Morgan RL, Kelley L, Guyatt G, Johnson A, Lavis JN. Decision-making frameworks and considerations for informing coverage decisions for healthcare interventions: a critical interpretive synthesis. *J Clin Epidemiol* 2018;94:143–50. [PubMed: 28988959]
- [38]. Kwon W, Clarke I, Wodak R. Organizational decision-making, discourse, and power: integrating across contexts and scales. *Discourse Commun* 2009;3(3):273–302.
- [39]. Martinovsky B. Discourse analysis of emotion in face-to-face group decision and negotiation In: *Emotion in Group Decision and Negotiation*. Netherlands: Springer Netherlands; 2015:137–88.



## What is new?

### Key findings

- Formal Grading of Recommendations Assessment, Development, and Evaluation (GRADE) criteria contributed to 94% of panel discussions (research evidence related to the clinical effects of a treatment or intervention [53%], resource use and costs [13%], balance of benefits and harms [13%], feasibility and acceptability [12%], patient values and preferences [3%], and health equity [2%]).
- Other GRADE-related factors (but not part of GRADE's formal criteria), contributed to 6% of panel discussions (clinical experience, political environment, and legal implications each occupied approximately 1%; how language should be used in guidelines represented 3%).
- The effect of instructions on strong normative rules (use of Evidence to Decision [EtD] framework to adhere to GRADE) may dominate other (non-GRADE) potential criteria that guidelines panelists may otherwise invoke.

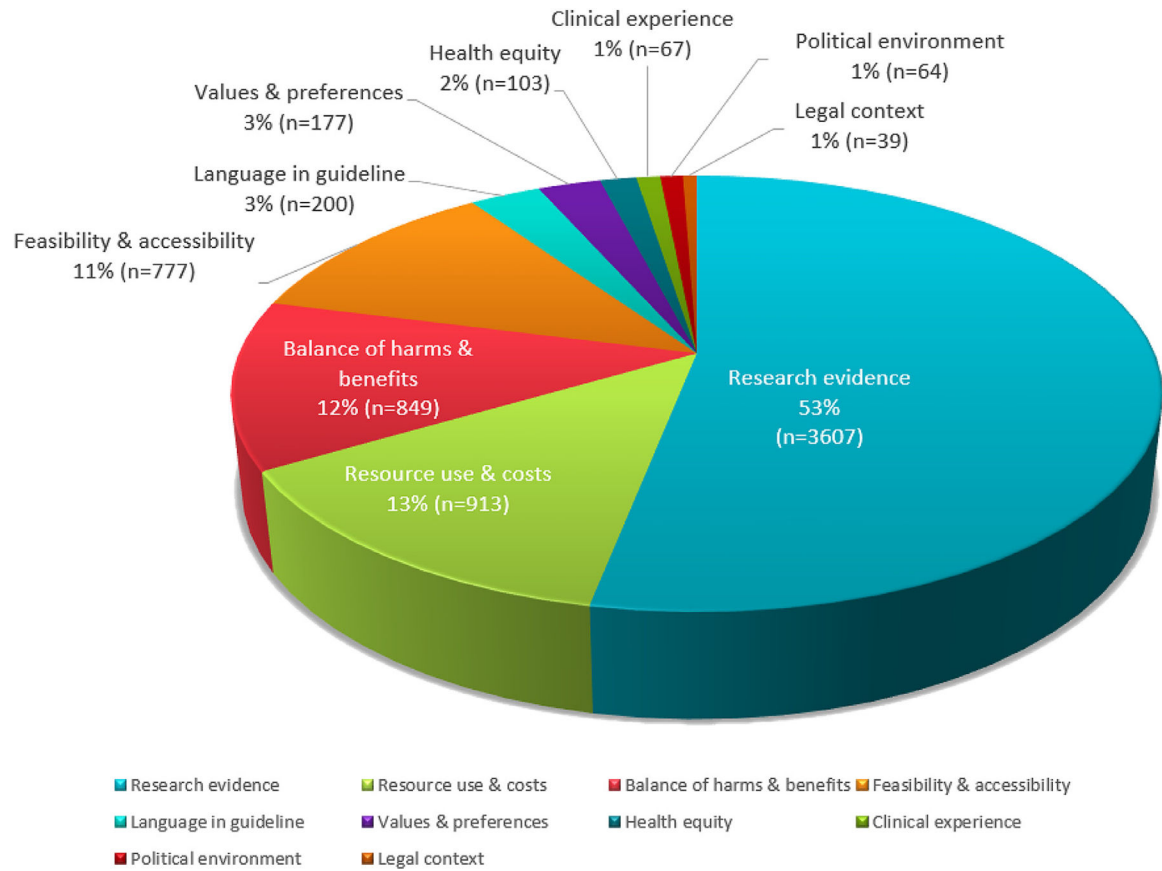
### What this adds to what was known?

- EtD frameworks provide structure for guidelines development meetings; however, panelists are likely to consider only those factors that GRADE mandates.
- Whether use of a highly structured format may un-intentionally repress discussion and consideration of domains outside of the EtD framework/ GRADE remains uncertain.
- It is also possible that the GRADE EtD framework is complete, and no other domains required consideration.

### What is the implication and what should change now?

- Guidelines developers should examine whether other factors (e.g., social context, individual characteristics of the panelists, cognitive biases, etc.) affect guidelines decision-making.
- Developers of GRADE and EtD can assess whether the proportion of discussions spent on each GRADE factor during guidelines panel meetings is appropriate.
- Researchers should compare real-time decision-making processes of groups that use the EtD framework to make GRADE guidelines recommendation with groups that do not.





**Fig. 1.** Display of percentage (rounded) and frequency of words or phrases deliberated by panelists for each theme.

**Table 1.**

Illustrative quotes for themes

| Themes                         | Supporting quotes  |
|--------------------------------|--|
| Research evidence              | <ul style="list-style-type: none"> <li>We would pool across different types of general surgeries and we would try to use that evidence or compare whether or not that evidence that is pooled, and that is somewhat indirect for this specific population, [but] is better than the evidence that we have here which is mainly plagued by imprecision.</li> <li>The concern is that you might just be pooling bad data, throwing bad data together, and you may have some good data in there that you throw together with bad data...so anyway.</li> <li>But the concept for now is we didn't specifically review this evidence, we didn't do the search, we didn't do the deconstruction; we used what the treatment for today and tomorrow the treatment panel gave us.</li> </ul> |
| Resource use and costs         | <ul style="list-style-type: none"> <li>So cost-effectiveness, meaning these assumptions about the events, does that then lead to spending the resources? Does that provide important benefit as opposed to just the cost.</li> <li>We can make recommendations or suggestions that assume a resource like an SRA is available and then also recommend that if that resource is not available.</li> </ul>   |
| Balance of benefits and harms  | <ul style="list-style-type: none"> <li>Inconsistencies, so do some of the study show benefit and some of the study show harm and there's differences in effect across the studies?</li> <li>In other words if you believe that the benefits don't clearly outweigh the risk or the risk don't clearly outweigh the benefits, or you think they are not relevant, that shifts our confidence levels for some of these outcomes.</li> <li>And...my physician made a decision [that] it was worth the risk when, you know the risk of bleeding if it's monitored correctly and you know what to look for, at least in my case, was worth it; was worth more than the harm of possibly thrombosis.</li> </ul>  |
| Feasibility and acceptability  | <ul style="list-style-type: none"> <li>Feasibility, acceptability and accessibility is okay; it may be similar in my country but when you go to other third world where it is been completely different.</li> <li>And acceptability, of course you may find something that is evidence-based, cost acceptable, feasible, and then no one wants to do it.</li> <li>Only consider an MRV [magnetic resonance venography] depending on availability because you can't say to people that you have to do and MRV when 50% of the countries can't get access.</li> </ul>  |
| Patient values and preferences | <ul style="list-style-type: none"> <li>I personally would say strong here just because of the level, the moderate evidence and the benefits in terms of acceptability from patient preferences standpoint.</li> <li>A personally I would say is almost a tie, so it can't be strongly supported but it can be left to patient values and preferences. I don't know.</li> <li>Could we say or would people feel comfortable, if we said we suggest against routine platelet count monitoring in medical patients receiving unfractionated heparin and then the values and preferences thing or do you feel like that is already too strong?</li> </ul>  |
| Health equity                  | <ul style="list-style-type: none"> <li>And these concepts are obviously closely tied to issues around feasibility, acceptability and equity.</li> <li>Health-equity though, in this case I think it holds because it means better compliance amongst patients with lower health literacy.</li> <li>The US healthcare system, if going to get into that say your treatment is going to be improved by testing and accost you 2,000 bucks and you don't have coverage for, that's going to increase health inequity.</li> </ul>  |
| Clinical experience            | <ul style="list-style-type: none"> <li>I have the gestalt this [situation] is not good then I am asking my best ultrasound clinician to go up really high. If the suspicion remains then sure, but so I instruct my patients that I think follow up point.</li> <li>So we can't pin down those risk factors and we are instead relying on gestalt of this surgeon. Probably we shouldn't be too prescriptive about it and have two specific recommendations.</li> </ul>  |

**Themes**

**Supporting quotes**

|                       |  |
|-----------------------|--|
|                       | <p>So generally we can either use what we have called in our discussions is the 'gestalt', where you would just say based on what I think clinically I think that this person is at risk of having HIT?</p>  |
| Political environment | <ul style="list-style-type: none"> <li>• And it might come at a cost but those costs are, those costs are of course caused by external people and if within this guideline you get an argument that we're recommending low molecular weight over unfractionated and that will help drive your political efforts to try and get the funding costs down because they'll say 'look at [what] ASH said, we should be using this'.</li> </ul> |
| Legal context         | <ul style="list-style-type: none"> <li>• These guidelines will have quite some medical legal implications and therefore I, we should clearly state that if the 4T score is properly performed and below four then maintaining, there is no, no reason to really change management.</li> </ul>  |
| Language in guideline | <ul style="list-style-type: none"> <li>• I think we all want to be as transparent as we can.</li> <li>• I think the bottom line is a long as we are explicit and sort of just say this is the rationale where we put it is a bit semantic.</li> <li>• But you are absolutely right; it needs to be transparently laid out that there are assumptions going into this.</li> </ul>   |