Journal of Evaluation in Clinical Practice

WILEY

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research

ORIGINAL PAPER OPEN ACCESS

Improving Guideline Development Processes: Integrating Evidence Estimation and Decision-Analytical Frameworks

Benjamin Djulbegovic[1] | Iztok Hozo[2] | Ilkka Kunnamo[3] | Gordon Guyatt[4]

[1]Department of Medicine, Division of Medical Hematology and Oncology, Medical University of South Carolina, Charleston, South Carolina, USA | [2]Department of Mathematics, Indiana University Northwest, Gary, Indiana, USA | [3]Duodecim Publishing Company, Helsinki, Finland | [4]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

Correspondence: Benjamin Djulbegovic (djulbegov@musc.edu)

Received: 19 January 2025 | **Revised:** 25 February 2025 | **Accepted:** 27 February 2025

Keywords: Bayesian statistics | decision analyses | decision-making | evidence-based medicine | frequentist statistics | GRADE | guidelines | inferences

ABSTRACT

Rationale, Aims and Objectives: Despite using state-of-the-art methodologies like Grades of Recommendation, Assessment, Development and Evaluation (GRADE), current guideline development frameworks still rely heavily on panellists' *intuitive* integration of evidence related to the benefits and harms/burdens of health interventions. This leads to the 'black-box' and 'integration' problems, highlighting the lack of transparency in guideline decision-making. Combined with humans' limited capacity to process the large volumes of information presented in Summary of Findings (SoF) tables—the primary output of systematic reviews that underpin guideline recommendations—this reliance on non-explicit processes raises concerns about the trustworthiness of clinical practice guidelines.

Methods: SoF tables provide the best available evidence, derived from frequentist or Bayesian estimation frameworks. Decision analysis, which integrates both types of estimates but considers intervention consequences, is the only analytical approach that combines multiple outcomes (benefits, harms and costs) into a single metric to support decision-making. Such analysis seeks to identify the optimal decision by balancing harms, benefits and uncertainties. This paper leverages the PICO format (Population, Intervention, Comparison(s), Outcome) as a conceptual basis for deriving SoF tables. Subsequently, we propose a solution to GRADE's "black-box" and "integration" problems by matching PICO-based SoF with decision models.

Results: We succeeded in connecting the PICO framework to simple decision-analytical models, restricted to time frames supported by empirically verifiable evidence, to calculate which competing intervention offers the greatest benefit (net differences in expected utility; ΔEU). The single metric [ΔEU] enabled a simple, transparent and easy-to-understand assessment of the superiority of competing management strategies across multiple outcomes (considering both benefits and harms), addressing the 'black-box' and 'integration' problems. Completing a SoF-based decision model takes about 10 min. Not surprisingly, the recommendations based on ΔEU may differ from the intuitive recommendations of panels.

Conclusion: We propose that incorporating the straightforward and transparent modelling into guideline panels' decision-making processes will enhance their intuitive judgements, resulting in more trustworthy recommendations. Given the simplicity of calculating ΔEU, we advocate for its immediate inclusion in systematic reviews and SoF tables.

Evidence-based medicine (EBM) has substantially contributed to the advancement of clinical practice [1]. Originally centred on critical appraisal and the development of systematic reviews (SRs), EBM, through the operationalization of its principles via the Grades of Recommendation Assessment, Development and Evaluation (GRADE) system for evaluating medical evidence, has increasingly contributed to advancements in methods for clinical practice guidelines (CPGs) [1, 2]. The latter has been considered a cornerstone for improving physicians' decision-making, and, in turn, patients' outcomes [1, 3]. Indeed, improving decision-making—poor decisions are considered the leading cause of death [4]—and inadequate adherence to CPGs—estimated to be the third leading cause of preventable patient deaths and one-third of unnecessary health care spending [5]—are central to rigorously developed evidence-based CPGs. The key difference between evidence-based CPGs and traditional consensus-based guidelines is that the former rely on SRs of the benefits and harms of different management options, but the latter do not [3]. This, in turn, requires matching the strength of recommendations (SoR) with the quality (certainty) of the supporting evidence (CoE) [6].

1 | A Brief Overview of the Development of Evidence-Based Guidelines

Evidence-based CPGs follow two steps. In the first step, SRs are conducted using the Population, Intervention, Comparison(s), Outcome, Timing and Setting (PICO (TS)) format to determine the scope of the evidence review and answer the questions posed [7]. The PICO format is widely regarded as one of the essential EBM tools for guiding the development of high-quality SRs [7–9]. The essence of PICO is a pair-wise comparison of competing management alternatives, restricted to one comparison at a time [7–9]. Other formats, such as SPIDER and SPICE [10], have been proposed for formulating research questions but are less promoted in the EBM and GRADE communities [11]. Our method aligns best with the PICO format, can be adapted to SPICE, but is less suited to SPIDER. Using a well-structured format to address one recommendation at a time helps enhance the scientific rigour of SRs and CPGs [7–9].

The end product of an SR is typically provided in meta-analytic estimates of the effects of health interventions summarized in the form of popular evidence and *SoF* tables [12]. These tables transparently present information about key outcomes (typically up to 7) related to the benefits and harms of competing health interventions, along with details about the quantity (number of studies and participants) and CoE, rated from very low to high [12, 13].

In the second step, a guidelines panel is formed, typically consisting of 10–20 individuals with both content and methodological expertise, as well as interest holders [14] representing patients and/or other parties identified as end-users of the guideline's recommendations [15]. The panel's task is to use the SoF tables to 'go from evidence to recommendations' [8, 16] and issue either strong or weak (conditional) recommendations in favour of one intervention over another.

Up to this point, the GRADE system is highly transparent, with well-defined inputs presented in the SoF tables—such as the CoE, benefits and harms, sometimes cost and assessment of patients' values and preferences (V&P)—and outputs expressed as the SoR for or against a particular health intervention.

However, how exactly panellists integrate the information provided in the SoF tables remains unclear. This has led to important criticism that GRADE suffers from a 'black-box' operation—a process with defined inputs and outputs but limited knowledge of its internal workings—and from the 'integration problem' referring to the lack of a theoretical framework of how CPG panel members should integrate evidence with other important factors such as benefits, harms and patients' V&P [17, 18]. Indeed, extending the GRADE system within decision-science platforms has been identified as a major challenge for advancing EBM and GRADE development over the next 25 years [1].

Here, we propose a solution to GRADE's 'black-box' and 'integration' problems by matching PICO-based SoF with decision models. Before providing conceptual details and concrete examples, let us address the key theoretical issues that explain why the current GRADE (and similar) methodology is not well suited to overcome these problems.

2 | Summary of Findings Tables Are Based on Estimates Derived From Both the Frequentist and Bayesian Statistical Frameworks

From an epistemological and statistical standpoint, EBM and GRADE have primarily relied on estimation within the frequentist framework, with occasional use of the Bayesian framework [19]. Operationally, this process culminates in assessing the estimated mean effects of interventions, with 95% confidence intervals for all outcomes of interest, as shown in the SoF tables (e.g., Table 1). Alternatively, it may involve estimating the posterior probability, with associated 95% credible intervals, that one treatment is better than another for each outcome.

In the frequentist framework, conclusions are based on statistical hypothesis testing, confidence intervals and statistical significance, typically related to one primary outcome [20]. This approach only considers whether there is enough evidence to reject the null hypothesis based on the observed data. In Bayesian analysis, the posterior probabilities of one treatment versus another are directly compared to assess treatment superiority.

Neither the frequentist nor the Bayesian frameworks explicitly account for the consequences of the interventions tested (e.g., benefits, harms, or costs). More importantly, neither framework can combine all treatment effects into a *single metric* to determine which intervention should be favoured. The integration of these elements is left to guideline panels, which *intuitively* combine them and subsequently make recommendations for or against interventions. However, the lack of explicitness and transparency, coupled with the human brain's limited capacity to integrate the large amount of information presented in SoF tables, raises concerns about the accuracy of the

TABLE 1 | An example of summary of findings table**.

Outcomes	Anticipated absolute effects* (95% CI) - risk with no parenteral anticoagulant	Anticipated absolute effects (95% CI) - risk with any parenteral anticoagulation (UFH, LMWH or fondaparinux)	Relative effect (95% CI)	No. of participants (studies)	Certainty of the evidence (GRADE) & comments
Mortality (all-cause)	69 per 1000	67 per 1000 (63 to 72)	RR 0.97 (0.91 to 1.04)	49,002 (21 RCTs)	LOW
Symptomatic pulmonary embolism	10 per 1000	6 per 1000 (5 to 8)	RR 0.59 (0.45 to 0.78)	25687 (13 RCTs)	MODERATE
Proximal deep vein thrombosis	4 per 1000	1 per 1000 (0 to 5)	RR 0.28 (0.06 to 1.37)	3706 (1 RCT)	MODERATE
Distal deep vein thrombosis	2 per 1000	2 per 1000 (0 to 7)	RR 0.75 (0.17 to 3.34)	3706 (1 RCT)	MODERATE
Major bleeding	7 per 1000	10 per 1000 (6 to 19)	RR 1.48 (0.81 to 2.71)	30,761(16 RCTs)	LOW
Gastrointestinal bleeding	31 per 1000	82 per 1000 (11 to 589)	RR 2.61 (0.36 to 18.86)	185 (2 RCTs)	LOW
Heparin-induced thrombocytopenia	2 per 1000	2 per 1000 (1 to 4)	RR 0.95 (0.47 to 1.92)	12,577 (3 RCTs)	MODERATE

Note: Any parenteral anticoagulation (UFH, LMWH or fondaparinux) compared to no parenteral anticoagulant in acutely ill medical patients for VTE prophylaxis.
***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
Patient or population: acutely ill medical patients for VTE prophylaxis; **Setting**: Inpatient; **Intervention**: any parenteral anticoagulation (UFH, LMWH or fondaparinux); **Comparison**: no parenteral anticoagulant
**see details at https://guidelines.gradepro.org/profile/54B577E9-7F80-3A78-B3EA-3850E9A1D432

recommendations. In this presentation, transparency in decision-making means clearly communicating the criteria, factors and values shaping a choice or policy, along with their relative importance. When guideline panels outline their decision-making process step by step, they enhance trust, clarity and accountability. This transparency underscores the need to quantify net effect estimates and assess their certainty, as discussed in a GRADE concept paper [21].

3 | Decision Analysis Takes the Consequences of Interventions Into Account

Decision analysis (DA) can integrate both frequentist and Bayesian estimates; it provides the *only* known analytical framework that combines multiple outcomes (both benefits and harms) into a *single metric* to aid decision-making [22]. It focuses on making the best decision by weighing the harms, benefits and uncertainties, thus accounting for the consequences of choosing one treatment over another. While there are many decision theories, the most used is the expected utility theory (EUT), which is widely employed in medical decision-making. EUT is the only theory that satisfies all mathematical and statistical axioms of rationality [22].

Classical rules of statistical inference do not apply to decision-analytical modelling [23]. Decisions should be based solely on the mean (or, median) expected utility (EU) metrics, regardless of whether the differences are statistically significant [23]. Box 1 illustrate comparison of the frequentist and decision-analytical approaches to making clinical recommendations.

DA has been criticized for employing complex procedures, such as microsimulation, Monte Carlo simulations, or Markov modelling, which can be opaque, non-transparent and based on empirically unverifiable assumptions [22]. On the other hand, current evidence-based guidelines restrict themselves to time frames supported by empirically verifiable evidence. In this article, we propose that the solution to the problem—the theoretical impossibility of aggregating all outcomes within the frequentist or Bayesian frameworks—lies in *using DA constrained by the empirical evidence* presented in SRs used by guidelines panels. As it turns out, many decision models are simple, transparent, easy to understand, and far superior to the existing 'black-box' guidelines process. Importantly, these decision models align directly with SoF tables and follow the PICO platform (Figure 1). Thus, if the panel has not considered certain elements in their recommendations, a DA should not include them either.

As noted by Tukey almost 65 years ago [24], we should distinguish 'conclusions' from 'decisions'. Both frequentist and Bayesian statistics are viewed as *estimation framework*s concerned with 'conclusions', that is, assessments of the truthfulness of findings under formalized inferential assumptions. On the other hand, DA is generally classified as a 'decision-making framework', focused on the *consequences* of specific actions in particular circumstances. Even when no clear 'winner' emerges from the estimation framework, we can still make an optimal decision based on the option that offers the highest mean EU, prioritizing long-term benefits over statistical significance or probability estimates.

BOX 1 | Comparison of the frequentist and decision-analytical approaches to making clinical recommendations: A hypothetical example.

Scenario: Consider a clinical trial comparing two treatments, *Treatment A* and *Treatment B*, for a certain medical condition. The outcomes of interest are:

1. *Mortality* (death within 1 year)
2. *Adverse Event 1* (severe allergic reaction)
3. *Adverse Event 2* (gastrointestinal complications)

Clinical Trial Results
The trial reports the following outcome rates:

- *Mortality Rates*:
    - Treatment A: 8%
    - Treatment B: 12%
    - *p-value*: 0.04
- *Adverse Event 1 Rates*:
    - Treatment A: 15%
    - Treatment B: 10%
    - *p-value*: 0.05
- *Adverse Event 2 Rates*:
    - Treatment A: 20%
    - Treatment B: 15%
    - *p-value*: 0.08

Frequentist Approach
In the frequentist framework, decisions are often based on statistical significance (commonly $p < 0.05$).

- *Mortality*: Treatment A shows a *significantly lower* mortality rate ($p = 0.04$)
- *Adverse Event 1*: Treatment A has a *significantly higher* rate of severe allergic reactions ($p = 0.05$).
- *Adverse Event 2*: The difference is *not statistically significant* ($p = 0.08$).

Recommendation/decision: Given the critical importance of mortality and its significant reduction with Treatment A, the frequentist approach would *favor Treatment A*.

Decision-Analytical Approach
The decision-analytical method considers both the probabilities of outcomes** and their relative importance. Here, we'll assign equal weights ($W = 1$) to all outcomes, reflecting that each is equally important.
Assigning Weights (using equal weights, $W = 1$)*:

- *Mortality (W_1): 1*
- *Adverse Event 1 (W_2): 1*
- *Adverse Event 2 (W_3): 1*

Calculating Weighted Adverse Outcome Scores (WAOS):
For each treatment, we calculate the WAOS:

$$WAOS = (W1 \times Mortality\ Rate) + (W2 \times Adverse\ Event_1\ Rate) + (W3 \times Adverse\ Event_2\ Rate)$$

Treatment A:

$$WAOSA = (1 \times 0.08) + (1 \times 0.15) + (1 \times 0.20) = 0.43$$

Treatment B:

$$WAOSB = (1 \times 0.12) + (1 \times 0.10) + (1 \times 0.15) = 0.37$$

Interpretation:
Treatment A has a *higher WAOS* (0.43) compared to *Treatment B* (0.37), indicating a worse overall profile when considering all outcomes equally.
Decision/recommendation: The decision-analytical approach, which aggregates all outcomes with equal importance, would *favor Treatment B* due to its *lower overall adverse outcome score*.

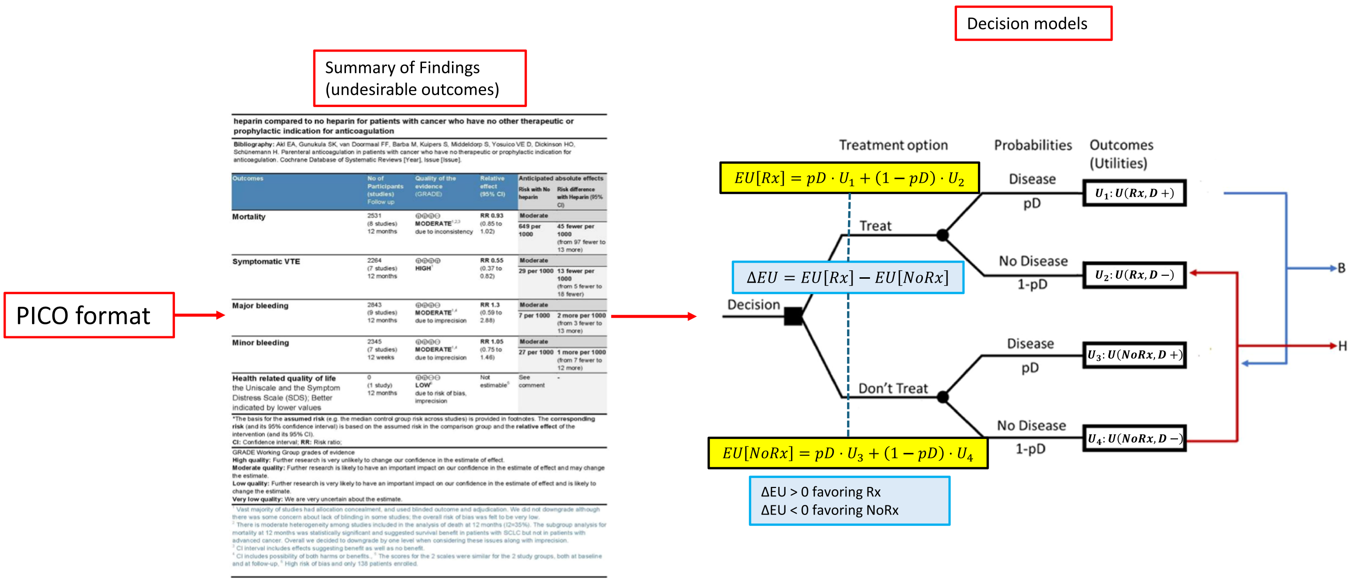
Conclusions

- *Frequentist Approach*: Focuses on individual outcomes and their statistical significance. In this case, the significant reduction in mortality with Treatment A leads to its preference.
- *Decision-Analytical Approach*: Considers the combined impact of all outcomes, weighed equally here. The higher rates of adverse events with Treatment A result in a higher overall adverse outcome score, leading to a preference for Treatment B.

This example illustrates how different analytical approaches can lead to contrasting treatment recommendations, especially when multiple outcomes with varying importance are considered.

W*–weights are referred to in the manuscript as relative values (RV) that capture patients' values and preferences (V&P) regarding the importance of avoiding one outcome over another. Note that, for most patients, it would make no sense to assign the same weight to mortality as to allergies, highlighting the importance of a decision-analytical over the frequentist approach (see Box 2 and the EXCEL calculators for evaluating the effect of RV on treatment recommendations).

** Note that in our method (see Section 4 conceptual Outline Appendix and EXCEL calculator), we assume that the probability of disease is often equal to 1. Under these circumstances, WAOS=1–expected utility (see also Box 2).



heparin compared to no heparin for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Bibliography: Akl EA, Gunukula SK, van Doormaal FF, Barba M, Kuipers S, Middeldorp S, Yosuico VE D, Dickinson HO, Schünemann H. Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation. Cochrane Database of Systematic Reviews [Year], Issue [Issue].

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No heparin	Risk difference with Heparin (95% CI)
Mortality	2531 (8 studies) 12 months	⊕⊕⊕⊝ **MODERATE**[1,2,3] due to inconsistency	**RR 0.93** (0.85 to 1.02)	**Moderate**	
				649 per 1000	**45 fewer per 1000** (from 97 fewer to 13 more)
Symptomatic VTE	2264 (7 studies) 12 months	⊕⊕⊕⊕ **HIGH**[1]	**RR 0.55** (0.37 to 0.82)	**Moderate**	
				29 per 1000	**13 fewer per 1000** (from 5 fewer to 18 fewer)
Major bleeding	2843 (9 studies) 12 months	⊕⊕⊕⊝ **MODERATE**[1,4] due to imprecision	**RR 1.3** (0.59 to 2.88)	**Moderate**	
				7 per 1000	**2 more per 1000** (from 3 fewer to 13 more)
Minor bleeding	2345 (7 studies) 12 weeks	⊕⊕⊕⊝ **MODERATE**[1,4] due to imprecision	**RR 1.05** (0.75 to 1.46)	**Moderate**	
				27 per 1000	**1 more per 1000** (from 7 fewer to 12 more)
Health related quality of life the Uniscale and the Symptom Distress Scale (SDS); Better indicated by lower values	0 (1 study) 12 months	⊕⊕⊝⊝ **LOW**[6] due to risk of bias, imprecision	Not estimable[5]	See comment	-

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.
[2] There is moderate heterogeneity among studies included in the analysis of death at 12 months (I2=35%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.
[3] CI interval includes effects suggesting benefit as well as no benefit.
[4] CI includes possibility of both harms or benefits., [5] The scores for the 2 scales were similar for the 2 study groups, both at baseline and at follow-up. [6] High risk of bias and only 138 patients enrolled.

FIGURE 1 | A conceptual outline of the model. The figure illustrates the use of the widely accepted PICO format—Population, Intervention, Comparison(s), Outcome—as the conceptual basis for deriving the Summary of Findings. This, in turn, is linked to simple decision-analytical models within a time frame supported by empirically verifiable evidence to determine which competing intervention is associated with the maximum benefit (net differences in expected utility; ΔEU). The calculation varies depending on whether the probability of disease (pD) is assumed to be equal to 1 or less than 1 (see the text and Appendix for details). U_1 to U_4 refer to the outcomes (utilities) of treatment (Rx) or no treatment (NoRx). D+ refers to outcomes when the disease is present; D− refers to outcomes when the disease is not present. B-net benefits; H-net harms. [Color figure can be viewed at wileyonlinelibrary.com]

4 | A Conceptual Outline

Figure 1 provides a conceptual outline of our proposal to link SoF tables (evidence estimates as inputs) with decision models leveraging the PICO format. As explained earlier, the SoF tables are populated with meta-analytical estimates of the competing interventions under consideration (e.g., treatment vs. no treatment, treatment #1 vs. treatment #2, etc.). Decision models also follow the PICO format.

The decision tree is solved using the 'folding-back' method (Appendix provides further details): outcomes are multiplied by the probability of disease (pD) to calculate the weighted average of outcomes, that is, the EU for each management alternative (e.g., EU for treatment [Rx] vs. EU for no treatment [NoRx]). The option with the highest EU is then selected as the best management choice. The model also considers the patient's V&P regarding different outcomes as illustrated in the example below.

The figure also defines the concept of net benefits (B) as the difference in the utility of outcomes when patients with the disease are treated with treatment #1 (Rx1) compared to not being treated (NoRx) or being treated with treatment #2 (Rx2). Net harms (H) are defined as the difference in the utility of outcomes when patients without the disease are not treated (NoRx) or treated with Rx2 compared to being treated with Rx1.

By solving the decision tree, we can derive two key assessments: (1) the overall best estimate, i.e., what is the best overall management strategy (by comparing EU between two treatments, ΔEU), and (2) individualized management recommendations by calculating a threshold for probability of disease (or a particular outcome) at which the EU of one treatment outweighs the EU of no treatment (or an alternative treatment). For simplicity, this paper focuses mainly on calculating the best strategy overall, as we believe that mandating the inclusion of EU for each management strategy in SoF tables would be most pragmatic step to improve the current intuitive approach to making guidelines recommendations. While determining the threshold (s) is also straightforward, it requires additional contextual details that make it challenging to automatically include in SoF tables.

A fundamental link between evidence-based guidelines and DA is the integration of evidence-based summary metrics—such as those included in SoF tables—either at the model's level of probabilities or utilities (outcomes) (see Appendix) [22, 25, 26]. For technical reason, to correctly solve decision models [22, 26], an SR team or guidelines panel should determine whether the diagnosis of a condition related to the outcomes listed in the SoF table is certain ($pD = 1$) or uncertain ($pD < 1$) [22, 26].

The clinical situations where the diagnosis is certain occur:

(a) when patients are enroled in a study after meeting specified diagnostic criteria, and

(b) when the diagnostic criteria are also part of the outcome definitions. This is common in fields such as cardiovascular diseases, haematology and oncology. For example, when modelling VTE recurrence, an established diagnosis requires the assumption of $pD = 1$ in the model [26] (Figure 1). A similar situation exists in oncology, where we often focus on detecting cancer relapses or metastasis (outcomes) in patients already diagnosed with cancer.

However, sometimes patients may be enroled in a study based on a diagnostic test, a predictive model, or a general assumption that the diagnosis is not certain. In such a decision model, we require the assumption of $pD < 1$.

While these scenarios are different, the structure of a decision model is identical (Figure 1), except for the differences in the assumptions of pD [$pD = 1$ vs. $pD < 1$], which generates different results.

- When $pD = 1$: because $pD = 1$, the lower branch in the tree becomes $(1 - pD) = 0$, then in this case $\Delta EU = B$ (see Figure 1), treatment with the more favourable utility is preferred (see the example below for further details):
- When $pD < 1$: $\Delta EU \neq B$; treatment with the more favourable EU is favoured.

In our experience, most SoF scenarios reflect the situation where $pD = 1$, which largely simplifies the mechanics of calculations and their interpretation.

5 | An Illustrative Example

American Society of Hematology (ASH) used the GRADE system to develop guidelines for prophylaxis of venous thromboembolism (VTE) [27]. One of the PICO they formulated was 'Should any parenteral anticoagulation (UFH, LMWH, or fondaparinux) versus no parenteral anticoagulant be used in acutely ill medical patients for VTE prophylaxis?' Table 1 shows the SoF table summarizing the effects of anticoagulation on eight outcomes: mortality, pulmonary embolism (PE), proximal distal vein thrombosis (DVT), distal DVT, overall major bleeding, gastro-intestinal (GI) major bleeding and heparin-induced thrombocytopenia (HIT). The patient's V&P regarding the importance of avoiding one outcome over another were expressed as relative values (RV), which reflect the preference (or weight) assigned to avoiding a VTE outcome or bleeding compared to mortality. Mortality is anchored at an RV of 100, representing the worst possible outcome. The RV for PE was assessed as 30, meaning it is 3.3 times (100/30) more desirable to avoid death than PE. Similarly, the RV for proximal DVT was 25, distal DVT 15, major bleeding 35, GI bleeding 15 and HIT 15. [For details on the various methods of eliciting V&P using this approach, please refer to the following reference [28, 29].]

Box 2 (and Appendix) shows the calculation of net benefits (ΔEU). In this case, we obtained that $\Delta EU = B < 0$ favoring NoRx. Note that differs from the ASH guidelines recommendations (Figure 2a).

Thus, from a DA point of view, the ASH recommendations seem not to be correct. Let's explore these differences in some details. First, during the deliberations, the panel may have considered other factors not displayed in the SoF tables. These include equity, acceptability and feasibility of implementing recommended interventions. However, in this case, these factors could not have played a role as a DA model recommended NoRx.

BOX 2 | An illustrative example: calculation of the net benefits of using anticoagulants for venous thromboembolism (VTE) prophylaxis.

American Society of Hematology (ASH) [ref # 27] panel made the following recommendations regarding VTE prophylaxis in acutely ill medical patients:

'In acutely ill medical patients, ASH guideline panel suggests using UFH, LMWH, or fondaparinux rather than no parenteral anticoagulant (conditional recommendation, low certainty in the evidence of effects)'.

To determine net benefits, we simply calculate the weighted sum of disutilities for outcomes in the presence of the disease, for treatment [Rx, D+] and no treatment [NoRx, D+] as listed in the Summary of Findings [SoF] Table 1 (see also Supporting Information S2: *App2.NetBenefit_instructions*):

Weighted Disutility (Rx, D+)

$$= RV_1 \cdot Mort + RV_2 \cdot PE + RV_3 \cdot ProxDVT + RV_4 \cdot DistDVT + RV_5 \cdot MajBleed + RV_6 \cdot GIBleed + RV_7 \cdot HIT$$

$$=1 \cdot 0.067 + 0.30 \cdot 0.06 + 0.25 \cdot 0.01 + 0.15 \cdot 0.002 + 0.35 \cdot 0.01 + 0.15 \cdot 0.082 + 0.15 \cdot 0.002$$

$$= 0.08545.$$

Weighted Disutility (NoRx, D+)

$$= RV_1 \cdot Mort + RV_2 \cdot PE + RV_3 \cdot ProxDVT + RV_4 \cdot DistDVT + RV_5 \cdot MajBleed + RV_6 \cdot GIBleed + RV_7 \cdot HIT$$

$$= 1 \cdot 0.069 + 0.30 \cdot 0.01 + 0.25 \cdot 0.004 + 0.15 \cdot 0.002 + 0.35 \cdot 0.007 + 0.15 \cdot 0.031 + 0.15 \cdot 0.002$$

$$= 0.0807.$$

Differences in expected utilities (ΔEU) are then

ΔEU = EU[Rx] − EU[NoRx] = U_1 − U_3

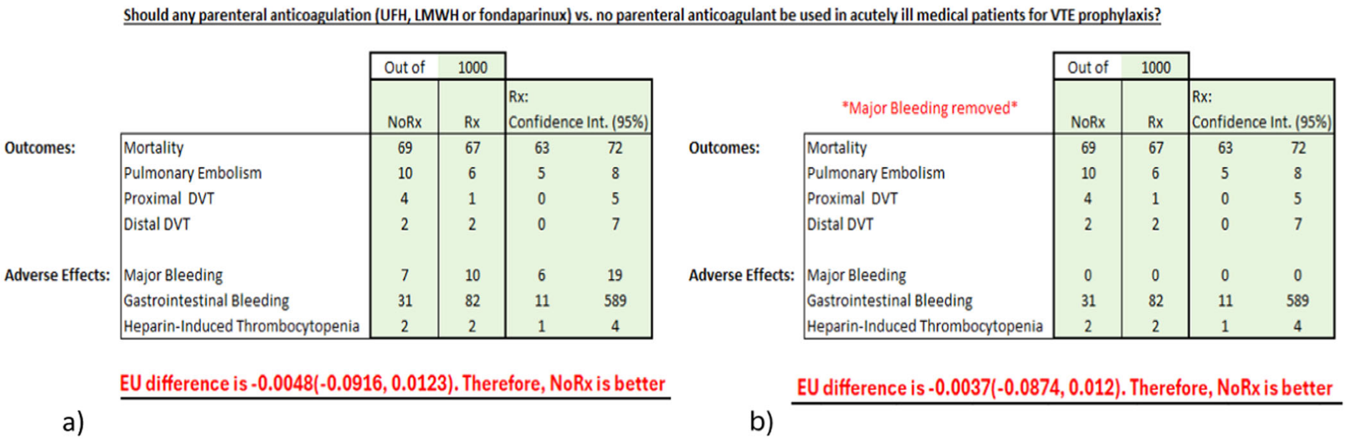
=(1 − Weighted Disutility (Rx, D+)) − (1 − Weighted Disutility (NoRx, D+))

=Weighted Disutility (NoRx) − Weighted Disutility (Rx)

$$=0.0807 - 0.08545 = -0.00475.$$

In this case, we obtained that $\Delta EU = B < 0$ favouring NoRx. Note that differs from the ASH guidelines recommendations (Figure 2a).

Second, although they present a summary of the best available evidence, almost all SoF tables have limitations, such as the potential for double counting. Currently, the only approach to address these limitations is to conduct a series of sensitivity analyses. For example, the table shows that the overall rate of major bleeding was lower than the GI bleed rate. This, of



Should any parenteral anticoagulation (UFH, LMWH or fondaparinux) vs. no parenteral anticoagulant be used in acutely ill medical patients for VTE prophylaxis?

a)

		Out of	1000		
		NoRx	Rx	Rx: Confidence Int. (95%)	
Outcomes:	Mortality	69	67	63	72
	Pulmonary Embolism	10	6	5	8
	Proximal DVT	4	1	0	5
	Distal DVT	2	2	0	7
Adverse Effects:	Major Bleeding	7	10	6	19
	Gastrointestinal Bleeding	31	82	11	589
	Heparin-Induced Thrombocytopenia	2	2	1	4

EU difference is -0.0048(-0.0916, 0.0123). Therefore, NoRx is better

b)

	Major Bleeding removed	Out of	1000		
		NoRx	Rx	Rx: Confidence Int. (95%)	
Outcomes:	Mortality	69	67	63	72
	Pulmonary Embolism	10	6	5	8
	Proximal DVT	4	1	0	5
	Distal DVT	2	2	0	7
Adverse Effects:	Major Bleeding	0	0	0	0
	Gastrointestinal Bleeding	31	82	11	589
	Heparin-Induced Thrombocytopenia	2	2	1	4

EU difference is -0.0037(-0.0874, 0.012). Therefore, NoRx is better

FIGURE 2 | Calculation of the net benefits of using anticoagulants for venous thromboembolism (VTE) prophylaxis (case when $pD = 1$). (a) Since the net benefits are negative ($B < 0$), no prophylaxis is preferred, although in extreme cases, prophylaxis may still be the better choice. (b) Sensitivity analysis when a potentially illogical entry (overall major bleeding rate) was excluded from the analysis. No prophylaxis remains the preferred choice, but note that in extreme cases, prophylaxis may still be a better option. pD, probability of disease. See the text, EXCEL calculator and Appendix for further details. [Color figure can be viewed at wileyonlinelibrary.com]

course, is logically implausible as the major GI bleed is a subset of the overall bleeding rate. Probably, this was a result of synthesizing completely different types of the studies after following the protocol for two separate PICOs for these two outcomes. Because GI bleed is often considered the most common and most serious harm, let's see what happens if we keep it in the analysis and drop the outcome for the overall bleeding rate.

As we can see, the results did not change much; $B = -0.0037$ still favours NoRx (Figure 2b). Do patients' V&P matter? Assuming (unrealistically) that V&P are equal across all outcomes (= 100), we also found that the results did not change. We can use the accompanying EXCEL file (Supporting Information S2: *App2.NB-instructions*) and change the assumptions to investigate the stability of our recommendations as we wish. For example, it is important to note that there is uncertainty in these assessments. While the best estimate favours NoRx ($B = -0.00475$), assuming a 'worst-case' scenario based on the upper 95% confidence values, B ranges from −0.0916 to 0.0123.

Because the SoF tables corresponding to $pD < 1$ is relatively rare, we show these calculations in the Appendix.

6 | Discussion

In this paper, we propose adding decision-analytical calculations of the overall benefits (expressed as Expected Utility, or EU) for each intervention to the SoF tables, along with ΔEU, to provide an initial assessment of the superiority of one intervention over another. Our proposal addresses three major issues with the current guidelines: the 'black-box' nature of decision-making, the 'integration problem' [17, 18] and the challenge of achieving a coherent theoretical integration between EBM and decision sciences [1]. We believe that *routinely* reporting information on ΔEU will be invaluable to guideline panels, as it will complement the currently non-transparent, intuitive methods of making recommendations with explicit and transparent guidance. When recommendations are neither transparent nor well-understood, they are more likely to be viewed as inaccurate.

We do not advocate ceding deliberative, considered judgements to formal models. Instead, we suggest using an explicit assessment of the utility of competing management alternatives as a starting point. Consider an intuitive estimation of, for example, multiplication of 34,567 * 23,587. Most people cannot accurately guess this product [815,331,829], but once they enter the numbers into a calculator, they will accept the result because they trust the mathematical principles of multiplication. According to the understanding/acceptance principle, the deeper the understanding of a normative principle, the more likely people are to accept the results and act accordingly [30].

Once a panel understands an unassailable logic supporting the use of decision theory, they will very likely accept its results to further guide their deliberation. While many decision-theoretical models exist [31], not all clinical circumstances may be captured within the outline shown in Figure 1 (although, we suspect, most will be). Additionally, EUT-based models do not explicitly capture emotions such as regret, which often influence decision-making [32, 33].

For this reason, we heed the time-honoured advice of cognitive scientists to 'value formal principles of rationality' [34], while also considering the need for adjustments that align with both explicit and implicit reasoning [34, 35]. Indeed, we previously argued that 'the most optimal decisions may be those that achieve coherence at both the normative and intuitive levels' [35]. Therefore, we recommend that guideline panels always begin by considering the EU values provided by the EUT-based model shown in Figure 1 and then adjust these values based on other elements (such as equity, acceptability, etc.) deemed essential for decision-making [21, 36].

The primary goal of this process is to contrast the panel's intuitive judgements with the model's explicit assessments. When the model disagrees with the panel's judgements, every effort should be made to explore, understand and reconcile these differences (as illustrated in the examples we discussed). We cannot overstate the importance of this exercise as the main pathway towards issuing reliable and trustworthy recommendations.

Importantly, end-users should have a clear understanding of how any adjustments—if made—were applied.

As we have emphasized throughout this paper, since our aim is to make an immediate impact on the development of CPGs, we have limited our proposal to the inclusion of EU assessments for pairwise PICO comparisons (see Figure 1). These calculations are straightforward—completing a SoF-based decision model takes about 10 min (see Appendix 3 Excel file for additional, randomly selected SoF-based calculations). They are much easier to understand than, for example, the statistical methods used in meta-analysis, which are widely accepted as the foundation for developing evidence-based guidelines. Based on our experience with CPG development, we believe that what we propose here will cover most clinical situations.

However, there is a growing number of network meta-analyses that assess comparisons of multiple interventions. Fortunately, the same decision model shown in Figure 1 applies to multiple comparisons but requires a different analytical framework—generalized decision curve analysis (gDCA)—which can be performed using both individual patient data [37] and aggregate data [38]. Similarly, the described models can be extended to handle diagnostic and prognostic/predictive assessments [37–39].

The primary strengths of these simple models are their transparency, explicitness and reliance on empirically verifiable assumptions (as provided through SoF tables). Nevertheless, in some cases, complex models—such as microsimulations, often based on many empirically unverifiable assumptions—may need to be used by guideline panels, as in the case of colorectal cancer screening guidelines [40]. As there is no evidence that complex models are empirically superior to simple models [22], panels must decide on a case-by-case basis when their SRs should be linked to simple versus complex models. For example, in cost-effectiveness analyses, a lifetime analytic horizon often requires complex models that extrapolate benefits, harms, or costs beyond available evidence. Conversely, limiting the analysis to empirical evidence typically applies only to short-term outcomes.

However, caution should be exercised regarding what is paradoxically considered a major methodological strength in developing evidence-based CPGs: the PICO format. Sometimes, biased judgements may arise from using pairwise comparisons when there are more than two options. For example, ASH guidelines for immune thrombocytopenia (ITP) [41] developed three recommendations for managing ITP in adults with no response to corticosteroids. By adhering to the pairwise PICO format, the panel first recommended either splenectomy or a thrombopoietin receptor agonists (TPO-RAs) (i.e., formally the panel judged splenectomy = TPO-RAs). In their second recommendation, the panel favoured the use of rituximab over splenectomy (i.e., formally rituximab > splenectomy). However, in the third comparison, the panel favoured administration of TPO-RAs rather than rituximab (i.e., TPO-RA > rituximab)! This sequence of recommendations violated the transitivity principle, one of the key axioms of rational decision-making. According to this principle, if we prefer A over B, and B over C, then we should rationally also prefer A over C [35].

BOX 3. | [Summary points].

The current methods for developing evidence-based clinical practice guidelines face several challenges:

- *Opaque evidence integration:* Recommendations often rely on panellists' *intuitive*, non-transparent integration of evidence on intervention benefits and harms. Critics describe these methods as a '*black-box*' process, with defined inputs and outputs but unclear internal mechanisms.
- *Lack of decision-theoretical framework:* Panellists typically base judgements on statistical evidence from frequentist or Bayesian frameworks, which do not account for the consequences of health interventions. This has been termed the '*integration problem*', highlighting the absence of a systematic approach to combine evidence with factors like benefits, harms and patient preferences.
- *Trustworthiness concerns:* The reliance on implicit processes undermines confidence in clinical practice guidelines.
- *Proposed solution:* We advocate for integrating decision analysis with frequentist and Bayesian estimates. Decision analysis is the *only* known analytical framework that can combine multiple outcomes (benefits and harms) into a single metric for informed decision-making.
- *Challenges in decision models:* Many decision models include assumptions that are opaque, non-transparent, or empirically unverifiable.
- *Recommended approach:* To address these issues, we propose decision analysis constrained by empirical evidence within the PICO framework (Population, Intervention, Comparison, Outcome), which guides systematic reviews and generates Summary of Findings (SoF) tables to inform panel deliberations and recommendations.
- *Calculation of net benefits:* By using simple decision-analytical models restricted to empirically supported time frames, we can calculate which intervention offers the greatest benefit (net differences in expected utility, ΔEU)—a crucial metric that is not explicitly included in contemporary guidelines
- *Utility of ΔEU:* We advocate for the immediate inclusion of ΔEU in systematic reviews and SoF tables. This metric is transparent, easy to calculate and will help enhance panellists' intuitive judgements, leading to more trustworthy recommendations.

A similar issue arose in recently published thrombophilia guidelines [42]. The panel faced a common clinical problem: after 3–6 months of treatment for a newly diagnosed VTE, they relied on pairwise comparisons between thrombophilia testing (A) and discontinuing anticoagulants (B), and, *separately*, between testing (A) and recommending indefinite anticoagulation for all patients (C). The panel failed to consider all relevant options *simultaneously* (A vs. B vs. C), leading to what is known as the 'omitted choice bias' [39]. When we compared the panel's guidelines against a decision model that accounted for all three management options, we found disagreement in about 48% (33/69) of the recommendations between the two approaches [39, 43].

These considerations highlight the risks of formulaic, automatic judgements. Interpreting scientific evidence and acting on it cannot be fully automated; it will always require careful, considered judgement, especially in the case of CPG recommendations that affect decision-making and health outcomes for thousands, if not millions, of people.

For this reason, we owe it to our patients to use the best possible methods to improve the accuracy of our recommendations. We argue that this can be achieved through closer integration of GRADE methodology with DA. Nevertheless, empirical testing is needed to determine how the proposed methods improve clinical guideline recommendations. The first step is a systematic evaluation of differences between panels using the GRADE approach with and without these methods. Box 3 provides a summary of the proposed method.

Acknowledgements

We thank Sydney Anuskiewicz for assisting in the creation of net benefit calculation examples shown in Appendix 2.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting material of this article.

References

1. B. Djulbegovic and G. H. Guyatt, “Progress in Evidence-Based Medicine: A Quarter Century On,” *Lancet* 390, no. 10092 (2017): 415–423, https://doi.org/10.1016/S0140-6736(16)31592-6.

2. G. H. Guyatt, A. D. Oxman, G. E. Vist, et al., “GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations,” *BMJ* 336, no. 7650 (2008): 924–926.

3. R. Graham, M. Mancher, D. M. Wolman, et al., *Clinical Practice Guidelines We Can Trust* (Washington, DC: Institute of Medicine, National Academies Press, 2011).

4. R. L. Keeney, “Personal Decisions Are the Leading Cause of Death,” *Operations Research* 56, no. 6 (2008): 1335–1347.

5. P. J. Pronovost, “Enhancing Physicians' Use of Clinical Guidelines,” *Journal of the American Medical Association* 310, no. 23 (2013): 2501–2502, https://doi.org/10.1001/jama.2013.281334.

6. B. Djulbegovic and G. Guyatt, “Evidence vs Consensus in Clinical Practice Guidelines,” *Journal of the American Medical Association* 322, no. 8 (2019): 725–726, https://doi.org/10.1001/jama.2019.9751.

7. J. Thomas, D. Kneale, J. E. McKenzie, et al., “Chapter 2: Determining the Scope of the Review and the Question It Will Address.” in *Cochrane Handbook of Systematic Reviews of Interventions, Version 64*, eds. J. P. T. Higgins, J. Thomas, J. Chandler, et al. (London: Cochrane, 2023).

8. G. H. Guyatt, A. D. Oxman, R. Kunz, et al., “Going From Evidence to Recommendations,” *BMJ* 336, no. 7652 (2008): 1049–1051.

9. IOM (Institute of Medicine), *Finding What Works in Health Care: Standards for Systematic Reviews* (Washington, DC: National Academies Press, 2011).

10. MUSC Libraries, *Question Formation Frameworks Charleston* (SC: Medical University of South Carolina, 2025), https://musc.libguides.com/systematicreviews/researchquestion.

11. A. M. Methley, S. Campbell, C. Chew-Graham, R. McNally, and S. Cheraghi-Sohi, “PICO, PICOS and SPIDER: A Comparison Study of Specificity and Sensitivity In Three Search Tools for Qualitative Systematic Reviews,” *BMC Health Services Research* 14, no. 1 (2014): 579, https://doi.org/10.1186/s12913-014-0579-0.

12. G. H. Guyatt, A. D. Oxman, N. Santesso, et al., “GRADE Guidelines: 12. Preparing Summary of Findings Tables-Binary Outcomes,” *Journal of Clinical Epidemiology* 66, no. 2 (2013): 158–172, https://doi.org/10.1016/j.jclinepi.2012.01.012.

13. N. Santesso, C. Glenton, P. Dahm, et al., “GRADE Guidelines 26: Informative Statements to Communicate the Findings of Systematic Reviews of Interventions,” *Journal of Clinical Epidemiology* 119 (2020): 126–135, https://doi.org/10.1016/j.jclinepi.2019.10.014.

14. N. A. Sathe, C. Ovelman, N. S. Ospina, et al., “Paper 6: Engaging Racially and Ethnically Diverse Interest Holders in Evidence Syntheses,” *Journal of Clinical Epidemiology* 176 (2024): 111575, https://doi.org/10.1016/j.jclinepi.2024.111575.

15. H. Schünemann, J. Brożek, G. Guyatt, et al., *The GRADE Working Group. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations* (Hamilton, Ontario, CA: McMaster University, 2013).

16. J. C. Andrews, H. J. Schünemann, A. D. Oxman, et al., “GRADE Guidelines 15: Going From Evidence to Recommendation-Determinants of a Recommendation's Direction and Strength,” *Journal of Clinical Epidemiology* 66, no. 7 (2013): 726–735, https://doi.org/10.1016/j.jclinepi.2013.02.003.

17. M. Mercuri, B. Baigrie, and R. E. G. Upshur, “Going From Evidence to Recommendations: Can GRADE Get Us There?,” *Journal of Evaluation in Clinical Practice* 24, no. 5 (2018): 1232–1239, https://doi.org/10.1111/jep.12857.

18. M. Mercuri and B. S. Baigrie, “What Confidence Should We Have in Grade?,” *Journal of Evaluation in Clinical Practice* 24, no. 5 (2018): 1240–1246, https://doi.org/10.1111/jep.12993.

19. B. Djulbegovic, G. H. Guyatt, and R. E. Ashcroft, “Epistemologic Inquiries in Evidence-Based Medicine,” *Cancer Control* 16, no. 2 (2009): 158–168.

20. P. Armitage, G. Berry, and J. N. S. Mathews, *Statistical Methods in Medical Research* (Oxford: Blackwell Science, 2002). 4th ed.

21. B. S. Alper, P. Oettgen, I. Kunnamo, et al., “Defining Certainty of Net Benefit: A GRADE Concept Paper,” *BMJ Open* 9, no. 6 (2019): e027445, https://doi.org/10.1136/bmjopen-2018-027445.

22. B. Djulbegovic and I. Hozo, *Threshold Decision-Making in Clinical Medicine. With Practical Application to Hematology and Oncology* (Cham, Switzerland: Springer Nature, 2023).

23. K. Claxton, “The Irrelevance of Inference: A Decision-Making Approach to the Stochastic Evaluation of Health Care Technologies,” *Journal of Health Economics* 18 (1999): 341–364.

24. J. W. Tukey, “Conclusions vs. Decisions,” *Technometrics* 2 (1960): 423–433.

25. B. Djulbegovic, I. Hozo, and G. H. Lyman, “Linking Evidence-Based Medicine Therapeutic Summary Measures to Clinical Decision Analysis,” *MedGenMed: Medscape General Medicine* 2, no. 1 (2000): 6.

26. B. Djulbegovic, I. Hozo, T. Mayrhofer, J. van den Ende, and G. Guyatt, “The Threshold Model Revisited,” *Journal of Evaluation in Clinical Practice* 25, no. 2 (2019): 186–195, https://doi.org/10.1111/jep.13091.

27. H. J. Schünemann, M. Cushman, A. E. Burnett, et al., “American Society of Hematology 2018 Guidelines for Management of Venous Thromboembolism: Prophylaxis for Hospitalized and Nonhospitalized

Medical Patients," *Blood Advances* 2, no. 22 (2018): 3198–3225, https://doi.org/10.1182/bloodadvances.2018022954.

28. B. Djulbegovic and I. Hozo, *Threshold Decision-Making in Clinical Medicine With Practical Application to Hematology and Oncology* (Cham, Switzerland: Springer Nature Switzerland AG, 2023).

29. B. Djulbegovic and I. Hozo, "Making Decisions When no Further Diagnostic Testing Is Available (Expected Regret Theory Threshold Model)," *Cancer Treatment and Research* 189 (2023): 39–52, https://doi.org/10.1007/978-3-031-37993-2_3.

30. P. Slovic and A. Tversky, "Who Accepts Savage's Axiom?," *Behavioral Science* 19, no. 6 (1974): 368–373, https://doi.org/10.1002/bs.3830190603.

31. L. He, W. J. Zhao, and S. Bhatia, "An Ontology of Decision Models," *Psychological Review* 129, no. 1 (2020): 49–72, https://doi.org/10.1037/rev0000231.

32. B. Djulbegovic and S. Elqayam, "Many Faces of Rationality: Implications of the Great Rationality Debate for Clinical Decision-Making," *Journal of Evaluation in Clinical Practice* 23, no. 5 (2017): 915–922, https://doi.org/10.1111/jep.12788.

33. B. Djulbegovic, I. Hozo, D. Lizarraga, and G. Guyatt, "Decomposing Clinical Practice Guidelines Panels' Deliberation Into Decision Theoretical Constructs," *Journal of Evaluation in Clinical Practice* 29, no. 3 (2023): 459–471, https://doi.org/10.1111/jep.13809.

34. K. E. Stanovich, "How to Think Rationally About World Problems," *Journal of Intelligence* 6, no. 2 (2018): 25, https://doi.org/10.3390/jintelligence6020025.

35. B. Djulbegovic and I. Hozo, "Evidence and Decision-Making," *Cancer Treatment and Research* 189 (2023): 1–24, https://doi.org/10.1007/978-3-031-37993-2_1.

36. B. Djulbegovic and I. Hozo, "Which Threshold Model?," *Cancer Treatment and Research* 189 (2023): 93–99, https://doi.org/10.1007/978-3-031-37993-2_8.

37. I. Hozo and B. Djulbegovic, "Generalised Decision Curve Analysis for Explicit Comparison of Treatment Effects," *Journal of Evaluation in Clinical Practice* 29, no. 8 (2023): 1271–1278, https://doi.org/10.1111/jep.13915.

38. I. Hozo, G. Guyatt, and B. Djulbegovic, "Decision Curve Analysis Based on Summary Data," *Journal of Evaluation in Clinical Practice* 30 (2024): 281–289, https://doi.org/10.1111/jep.13945.

39. B. Djulbegovic, I. Hozo, and G. H. Guyatt, "Decision Theoretical Foundation of Clinical Practice Guidelines: An Extension of the ASH Thrombophilia Guidelines," *Blood Advances* 8 (2024): 3596–3606, https://doi.org/10.1182/bloodadvances.2024012931.

40. L. M. Helsingen, P. O. Vandvik, H. C. Jodal, et al., "Colorectal Cancer Screening With Faecal Immunochemical Testing, Sigmoidoscopy or Colonoscopy: A Clinical Practice Guideline," *BMJ* 367 (2019): l5515, https://doi.org/10.1136/bmj.l5515.

41. C. Neunert, D. R. Terrell, D. M. Arnold, et al., "American Society of Hematology 2019 Guidelines for Immune Thrombocytopenia," *Blood Advances* 3, no. 23 (2019): 3829–3866, https://doi.org/10.1182/bloodadvances.2019000966.

42. S. Middeldorp, R. Nieuwlaat, L. Baumann Kreuziger, et al., "American Society of Hematology 2023 Guidelines for Management of Venous Thromboembolism: Thrombophilia Testing," *Blood Advances* 7, no. 22 (2023): 7101–7138, https://doi.org/10.1182/bloodadvances.2023010177.

43. I. Hozo and B. Djulbegovic, "Thrombophilia Management Calculator," *Blood Advances* 8, no. 15 (2024): 3914–3916, https://doi.org/10.1182/bloodadvances.2024013463.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.