Methodological Issues in Rating Certainty of Evidence and Interpreting Magnitude of Effect in Systematic Reviews and Practice Guidelines

By Linan Zeng

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment to the Requirements for the Degree Doctor of Philosophy

McMaster University (Health Research Methodology), Hamilton, Ontario

Doctor of Philosophy (2023)
Title: Methodological Issues in Rating Certainty of Evidence and Interpreting Magnitude of Effect in Systematic Reviews and Practice Guidelines

Author: Linan Zeng, Ph.D.
Supervisor: Romina Brignardello-Petersen D.D.S., M.S_C., Ph.D.
Number of pages: 165

Abstract

In the development of a *BMJ Rapid recommendation* – an international practice guideline initiative led by the *MAGIC Evidence Ecosystem Foundation*, and aiming to produce trustworthy, accessible and timely guidance – of plasma exchange and dosage of corticosteroids for patients with ANCA-associated vasculitis (AAV) (Chapter 2) two methodological issues arose.

The first issue is related to the rating of the certainty of evidence supporting the recommendations. Reviewers experienced challenges in making an explicit statement about what it was in which they were rating their certainty (i.e., the target of the rating of certainty of evidence). Through iterative discussions and presentations at *GRADE (Grading of Recommendations, Assessment, Development, and Evaluations) Working Group* meetings, the research team developed new GRADE guidance (Chapter 3 and 4) to help systematic reviewers be aware of the importance of determining the target of their rating of certainty of evidence and provided practical principles to help systematic reviewers specify this target.

The second issue arose from the process of moving from evidence to decisions. To help the *BMJ Rapid recommendation* panel interpret the magnitude of benefit and harm associated with plasma exchange, which required understanding patient values and preferences, the research team created a panel survey for eliciting the panelists' view regarding patient values and preferences. The research team then applied the panel survey approach in some other guidelines. Based on the experience of developing panel surveys, and through iterative discussions and consensus, the research team developed a framework for using surveys to guide guideline panels in making inferences regarding patient values and preferences (Chapter 5). Using interpretive description, the team conducted a qualitative evaluation regarding the influence of the panel surveys on the panels' understanding of patient values and preferences, interpretation of magnitude of benefits and harms, and on panels' decision on guideline recommendations (Chapter 6). The panel surveys proved to help guideline panels explicitly consider and incorporate patient values and preferences in making recommendations.

Acknowledgements

This work would not have been possible without the contribution and support of many people. I would like to express my gratitude to them.

Firstly, my supervisor, Dr. Romina Brignardello-Petersen, who I feel indebted for her mentorship, patience, and trust. Thank you, Romina, for allowing freedom to explore the research area I am interested.

My thesis committee members, Dr. Gordon Guyatt and Dr. Thomas Agoritsas, who provided me with generous guidance in navigating through the exciting world of methodological studies. Thank you Gordon for bringing me into "the deeper end of the pool", and your mentorship will inspire me throughout my career. Thank you Thomas for providing me with insightful feedback and smart advice whenever I needed it.

My colleagues Dr. Per O. Vandvik, Reed A.C. Siemieniuk and Lyubov Lytvyn at MAGIC. Thank you Per for welcoming me into the MAGIC family and always being supportive. Thanks Reed and Lyuba for navigating me in my first guideline project.

My colleagues at GRADE working group who supported with me in the GRADE methods projects particularly Dr. Monica Hultcrantz, Dr. Reem A. Mustafa, Dr. Mohammad H. Murad, Dr. Alfonso Iorio, Dr. Gregory Traversy, Dr. Ariel Izcovich, Dr. Hans de Beer and Dr. David Tovey.

I would like to thank all the participants in the interviews (especially Dr. Lise M. Helsingen and Dr. Nicolas Delvaux who participated in more than one interviews) for the time you put in this project and the invaluable feedback you gave me.

I would like to thank all staff in the Department of Health Research Methods, Evidence, and Impact for enhancing experience in the last few years.

And thank you most of all, my supportive family!

Table of contents

Lists of tables, figures, and appendices

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Abbreviations

1	AAV	ANCA-Associated Vasculitis
2	ANCA	Antineutrophil Cytoplasmic Antibody
3	ARCH	Arthritis Research and Collaboration Hub
4	ASFA	American Society For Apheresis
5	BMJ	British Medical Journal
6	BR	Baseline Risk
7	BSR	Brazilian Society of Rheumatology
8	BSR/BHPR	British Society for Rheumatology/British Health Professionals in Rheumatology
9	CanVasc	Canadian Vasculitis Research Network
10	CI	Confidence Interval
11	COREQ	Consolidated Criteria for Reporting Qualitative Research
12	CR	Colorectal
13	CRG	Cornerstone Research Group
14	ESKD	Developing End Stage Kidney Disease/End-Stage Kidney Disease
15	EULAR/ERA-EDTA	European League Against Rheumatism/European Renal Association-European Dialysis And Transplant Association
16	EVT	Endovascular Thrombectomy
17	GFR	Glomerular Filtration Rate
18	GI	Gastrointestinal
19	GIN	Guidelines International Network
20	GLP-1	Glucagon-Like Peptide 1
21	GPA	Granulomatosis With Polyangiitis
22	GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
23	HiREB	Hamilton Integrated Research Ethics Board
24	IGO	Improving Global Outcomes
25	ILSL	International Life Sciences Institute
26	IOM	Institute of Medicine
27	IVH	Intraventricular Haemorrhage
28	LDL	Low-Density Lipoprotein
29	MAGIC	The MAGIC Evidence Ecosystem Foundation
30	MID	Minimally Important Difference
31	MPA	Microscopic Polyangiitis
32	NICE	National Institute for Clinical Excellence
33	OR	Odds Ratio
34	PCSK	Proprotein Convertase Subtilisin/Kexin
35	PCSK-9	Proprotein Convertase Subtilisin/Kexin-9
36	PICOT	Patient, Intervention, Comparison, Outcomes and Timeline for Measuring the Outcomes
37	PLEX	Plasma Exchange
38	PPIs	Proton Pump Inhibitors

39	RCT	Randomised Controlled Trial
40	RD	Risk Difference
41	RR	Risk Ratio
42	SAPS	Subacromial Pain Syndrome
43	SGLT-2	Sodium-Glucose Co-Transporters-2
44	SGLT-2	Sodium-Glucose Transport Protein 2
45	VAS	Visual Analogue Scale
46	WHO	World Health Organization

Chapter 1

Introduction of the thesis

Clinical practice guidelines are statements that include recommendations intended to optimize patient care[1]. Until the 1970s, guidelines were primarily based on the consensus of experts[2]. Experts recommended management approaches they had used and, without an explicit systematic search, cited references they recalled or were able to identify [3]. With the emergence of a new method for physicians reading medical literature (i.e., critical appraisal) in the 1980s[4] and the introduction of evidence-based medicine as a principle for decision making in 1991[5], more rigorous approaches for guideline development have emerged.

The process for developing evidence-based guidelines includes defining the scope of guideline, formulating the questions and synthesizing the evidence, rating the quality of evidence (that reflects one's certainty that the true effect of an intervention falls on one side or another of a threshold or within a range)[6], and formulating the direction and strength of recommendations (the extent to which one can be confident that adherence will do more good than harm[6]) based on the quality of evidence, balance of benefits and harms, patient values and preferences, available resources, feasibility of the intervention, acceptability by stakeholders and impact on health equity[7].

Although the process for developing evidence-based guidelines is widely accepted, new methodological issues keep emerging. This thesis starts with the development of a *BMJ Rapid recommendation* – an international clinical practice guideline initiative founded and led by the *MAGIC Evidence Ecosystem Foundation*[8] aiming to produce trustworthy, accessible and timely guidance (**Chapter 2**). This *BMJ Rapid recommendation* of plasma exchange and dosage of corticosteroids for patients with ANCA-associated vasculitis (AAV), triggered by the publication of PEXIVAS randomised controlled trial (RCT), followed the process for developing evidence-based guidelines[9].

Two methodological issues arose when developing this guideline. First, authors of a linked systematic review discussed with methodologists on the guideline panel: when rating certainty of evidence using GRADE (Grading of Recommendations, Assessment, Development, and Evaluation), how to be clear about what it is in which they were rating their certainty of evidence. For example, in patients with baseline serum creatinine at 200-300 μmol/l, the risk difference in end-stage kidney disease between plasma exchange and usual care is 28 fewer events per 1000 patients with AAV with a 95% confidence interval from 43 fewer to 7 fewer per 1000 patients. Some may rate their certainty that compared to usual care plasma exchange reduces end-stage kidney disease and not rate down for imprecision (as the confidence interval did not cross the null effect threshold). Had the systematic reviewers have no concerns in the other GRADE domains (i.e., risk of bias, indirectness, inconsistency and publication bias), they would conclude that plasma exchange reduces end-stage kidney disease. Others might rate their certainty that there is an important difference in end-stage kidney disease between plasma exchange and usual care. If they set a minimally important difference (MID) at 10 fewer end-stage kidney disease per 1000 patients, they would rate down for imprecision (as the confidence interval crossed the MID threshold) and conclude that plasma exchange probably has an important reduction on end-stage kidney disease. Others may believe that the guideline panel would be optimally informed by seeing both ratings.

The study in **Chapter 3**, based on GRADE clarification of the construction of certainty of evidence in 2017[10], developed new GRADE guidance that provides practical principles and examples to help systematic review and health technology assessment authors make optimal choices regarding the target of their rating of certainty of evidence. The study in **Chapter 4** further discusses the choice of target of certainty of evidence rating in a specific situation; that is, when authors initially target the null effect threshold, and the point estimate is very close to the null, how they can decide whether to change the target of certainty rating or use the initial target.

The second methodological issue arose in the process of moving from evidence to recommendations. The systematic review revealed that the key benefit associated with plasma exchange is the reduction in end-stage kidney disease while the key harm is the increase in serious infections[11]. To interpret the magnitude of the key benefit and harm and to trade off the benefit and harm, the guideline panel need to estimate how much weight typical well-informed patients place on each of the two key outcomes.

Ideally, cross-sectional surveys among large samples of target patients would inform the relative importance that patients place on the outcomes. Unfortunately, such surveys among relevant patients were not available (which is often the case for guidelines[12]). Even when available, survey results often differ, and guideline panels need to interpret patient values and preferences based on the available evidence and their experience[13].

The research reported in **Chapter 5** introduced a novel framework for developing guideline panel surveys to help guideline panelists make inferences regarding patient values and preferences, interpret the magnitude of benefit and harm, and trade off the benefit and the harm associated with interventions under consideration. The research team developed this framework based on experience of applying the panel surveys in this *BMJ Rapid recommendation*[14] and in other guidelines[15-21]. The study reported in **Chapter 6** qualitatively evaluated the influence of the panel surveys on the panels' understanding of patient values and preferences, the panels' trade-off between benefits and harms with interventions, and the panels' decisions on the direction and strength of guideline recommendations.

Lastly, in **Chapter 7** we discuss the main findings and suggest directions for further research.

Reference

1. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, eds. *Clinical Practice Guidelines We Can Trust*. Washington (DC): National Academies Press (US); 2011.
2. Woolf SH. Practice guidelines, a new reality in medicine. II. Methods of developing guidelines. *Arch Intern Med*. 1992;152(5):946-952.
3. Murad MH. Clinical Practice Guidelines: A Primer on Development and Dissemination. *Mayo Clin Proc*. 2017;92(3):423-433. doi:10.1016/j.mayocp.2017.01.001
4. How to read clinical journals: I. why to read them and how to start reading them critically. Can Med Assoc J. 1981;124(5):555-558.

5. Guyatt G. Evidence-based medicine.*ACP J Club*. 1991; 114: A16 https://www.jameslindlibrary.org/guyatt-gh-1991/
6. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490. doi:10.1136/bmj.328.7454.1490
7. World Health Organization(WHO). WHO handbook for guideline development, 2nd Edition. https://www.who.int/publications/i/item/9789241548960
8. MAGIC (Making Grade the Irresistible Choice). https://magicevidence.org
9. Walsh M, Merkel PA, Peh CA, et al. Plasma Exchange and Glucocorticoids in Severe ANCA-Associated Vasculitis. *N Engl J Med*. 2020;382(7):622-631. doi:10.1056/NEJMoa1803537
10. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006
11. Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: an updated systematic review and meta-analysis. *BMJ*. 2022;376:e064604. doi:10.1136/bmj-2021-064604
12. Kim C, Berta WB, Gagliardi AR. Exploring approaches to identify, incorporate and report patient preferences in clinical guidelines: Qualitative interviews with guideline developers. *Patient Educ Couns*. 2021;104(4):703-708. doi:10.1016/j.pec.2020.10.001
13. Blackwood J, Armstrong MJ, Schaefer C, et al. How do guideline developers identify, incorporate and report patient preferences? An international cross-sectional survey. *BMC Health Serv Res*. 2020;20(1):458. doi:10.1186/s12913-020-05343-x
14. Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ*. 2022;376:e064597. doi:10.1136/bmj-2021-064597
15. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;367:l5515. doi:10.1136/bmj.l5515
16. Word Health Organization. Therapeutics and COVID-19: living guideline, https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-therapeutics-2022.2; 2022
17. Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ*. 2021;373:n1091. doi:10.1136/bmj.n1091
18. Hao Q, Aertgeerts B, Guyatt G, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations. *BMJ*. 2022;377:e069066. doi:10.1136/bmj-2021-069066
19. Ye Z, Busse JW, Hill MD, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: A clinical practice guideline. *J Evid Based Med*. 2022;15(3):263-271. doi:10.1111/jebm.12493
20. Conservative, pharmacological and surgical interventions for patients with temporomandibular disorder: a clinical practice guideline. 2023 [Forthcoming].
21. Interventional procedures for axial and radicular chronic spine pain: a clinical practice guideline. 2023 [Forthcoming].

Chapter 2

Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline

Zeng L, Walsh M, Guyatt GH, Siemieniuk RAC, Collister D, Booth M, Brown P, Farrar L, Farrar M, Firth T, Fussner LA, Kilian K, Little MA, Mavrakanas TA, Mustafa RA, Piram M, Stamp LK, Xiao Y, Lytvyn L, Agoritsas T, Vandvik PO, Mahr A

The BMJ; 2022; 376:e064597

Abstract

Clinical questions What is the role of plasma exchange and what is the optimal dose of glucocorticoids in the first 6 months of therapy of patients with antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV)? This guideline was triggered by the publication of a new randomised controlled trial.

Current practice Existing guideline recommendations vary regarding the use of plasma exchange in AAV and lack explicit recommendations regarding the tapering regimen of glucocorticoids during induction therapy.

Recommendations The guideline panel makes a weak recommendation against plasma exchange in patients with low or low-moderate risk of developing end stage kidney disease (ESKD), and a weak recommendation in favour of plasma exchange in patients with moderate-high or high risk of developing ESKD. For patients with pulmonary haemorrhage without renal involvement, the panel suggests not using plasma exchange (weak recommendation). The panel made a strong recommendation in favour of a reduced dose rather than standard dose regimen of glucocorticoids, which involves a more rapid taper rate and lower cumulative dose during the first six months of therapy.

How this guideline was created A guideline panel including patients, a care giver, clinicians, content experts, and methodologists produced these recommendations using GRADE and in adherence with standards for trustworthy guidelines. The recommendations are based on two linked systematic reviews. The panel took an individual patient perspective in the development of recommendations.

The evidence The systematic review of plasma exchange identified nine randomised controlled trials (RCTs) that enrolled 1060 patients with AAV. Plasma exchange probably has little or no effect on mortality or disease relapse (moderate and low certainty). Plasma exchange probably reduces the one year risk of ESKD (approximately 0.1% reduction in those with low risk, 2.1% reduction in those with low-moderate risk, 4.6% reduction in those with moderate-high risk, and 16.0% reduction in those with high risk or requiring dialysis) but increases the risk of serious infections (approximately 2.7% increase in those with low risk, 4.9% increase in those with low-moderate risk, 8.5% increase in those with moderate-high risk, to 13.5% in high risk group) at 1 year (moderate to high certainty). The guideline panel agreed that most patients with low or low-moderate risk of developing ESKD would consider the harms to outweigh the benefits, while most of those with moderate-high or high risk would consider the benefits to outweigh the harms. For patients with pulmonary haemorrhage without kidney involvement, based on indirect evidence, plasma exchange may have little or no effect on death (very low certainty) but may have an important increase in serious infections at 1 year (approximately 6.8% increase, low certainty). The systematic review of different dose regimens of glucocorticoids identified two RCTs at low risk of bias with 704 and 140 patients respectively. A reduced dose regimen of glucocorticoid probably reduces the risk of serious infections by approximately 5.9% to 12.8% and probably does not increase the risk of ESKD at the follow-up of 6 months to longer than 1 year (moderate certainty for both outcomes).

Understanding the recommendation The recommendations were made with the understanding that patients would place a high value on reduction in ESKD and less value on avoiding serious infections. The panel concluded that most (50-90%) of fully informed patients with AAV and with low or low-moderate risk of developing ESKD with or without pulmonary haemorrhage would

decline plasma exchange, whereas most patients with moderate-high or high risk or requiring dialysis with or without pulmonary haemorrhage would choose to receive plasma exchange. The panel also inferred that the majority of fully informed patients with pulmonary haemorrhage without kidney involvement would decline plasma exchange and that all or almost all (≥90%) fully informed patients with AAV would choose a reduced dose regimen of glucocorticoids during the first 6 months of therapy.

This *BMJ Rapid Recommendations* article is one of a series that provides clinicians with trustworthy recommendations for potentially practice changing evidence. *BMJ Rapid Recommendations* represent a collaborative effort between the *MAGIC* group (www.magicevidence.org) and *The BMJ*. A summary is offered here, and the full version including decision aids is on the MAGICapp (www.magicapp.org), for all devices in multilayered formats. Those reading and using these recommendations should consider individual patient circumstances and their values and preferences and may want to use consultation decision aids in MAGICapp to facilitate shared decision making with patients. We encourage adaptation of recommendations to allow contextualisation of recommendations and to reduce duplication of work. Those considering use or adaptation of content may go to MAGICapp to link or extract its content or contact *The BMJ* for permission to reuse content in this article.

Antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV), which includes granulomatosis with polyangiitis and microscopic polyangiitis, is characterised by inflammation of small blood vessels (see box 1 for details of AAV)[1]. Over the past few decades, with the evolution of disease awareness, diagnostic techniques, and improved treatments, mortality among patients with AAV has decreased[2]. However, it remains 2.6-fold higher than that in the general population due to complications from the underlying disease (such as kidney failure or pulmonary haemorrhage) and complications from immunosuppressive therapy (such as serious infections or cancer)[3,4].

Box Start

Box 1 Details of antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis

Classification—ANCA-associated vasculitis (AAV) includes granulomatosis with polyangiitis (GPA), and microscopic polyangiitis (MPA)[25]. An alternative approach to classification is based on ANCA serology (myeloperoxidase ANCA or proteinase 3 ANCA).

Clinical presentation—Typical features of GPA include nasal crusting, stuffiness, and epistaxis; scleritis; upper respiratory tract involvement; and often, when in the context of an active urinary sediment, kidney involvement. Patients with MPA are typically older and present with more severe kidney disease than those with GPA[26]. All forms of AAV can involve pulmonary haemorrhage.

Pathophysiology—Because both myeloperoxidase and proteinase 3 are sequestered from the immune system in primary granules and, after neutrophil degranulation at sites of tissue injury, are rapidly eliminated by specific inhibitors, it is unclear why autoantibodies to neutrophil self antigens develop. Defective neutrophil apoptosis or impaired clearance of apoptotic cell fragments may lead to prolonged exposure of the immune system to these antigens. Infection may also play a role through molecular mimicry[25].

Treatment—Initial therapy for AAV includes induction of remission with initial immunosuppressive therapy (treatment options include glucocorticoids, rituximab, cyclophosphamide, C5a inhibitors, mycophenolate mofetil, plasma exchange, intravenous immunoglobulin, and co-trimoxazole), and maintenance of remission with immunosuppressive therapy for a variable period to prevent relapse (treatment options include glucocorticoids, azathioprine, methotrexate, rituximab, and co-trimoxazole)[30].

Box End

This clinical practice guideline was triggered by publication of the PEXIVAS randomised controlled trial (RCT), holding the potential to change clinical practice[5]. The PEXIVAS trial failed to show a reduction in the composite outcome of death from any cause or end stage kidney disease (ESKD) in patients with severe AAV (defined by an estimated glomerular filtration rate (eGFR) of <50 mL/min/1.73 m^2 of body surface area or diffuse pulmonary haemorrhage) randomised to plasma exchange in addition to immunosuppressive therapy compared with immunosuppressive therapy alone (28.4% *v* 31.0%, hazard ratio 0.86 (95% confidence interval (CI) 0.65 to 1.13))[5]. This trial demonstrated that a reduced dose regimen of glucocorticoid (the cumulative dose was 40% of that in a standard dose regimen group at 6 months) reduced serious infections at 1 year compared with the standard dose regimen group (incidence rate ratio 0.69 (95% CI 0.52 to 0.93)).

We translated this new evidence for clinicians and patients using the GRADE approach and standards for trustworthy guidelines, as for previous *BMJ Rapid Recommendations* (see box 2). The guideline panel asked three key questions:

1. Which patients with AAV and kidney involvement, if any, should receive plasma exchange?
2. Should patients with AAV and pulmonary haemorrhage without kidney involvement receive plasma exchange?
3. Should patients with AAV receive a reduced dose regimen of glucocorticoid during the first 6 months of therapy?

Box Start

Box 2 How these recommendations were developed

The *BMJ Rapid Recommendations* was initiated by the *MAGIC Evidence Ecosystem Foundation* (MAGIC, https://magicevidence.org/) together with *The BMJ* in 2016 to circumvent organisational barriers and to provide clinicians with guidance based on the most current practice-changing evidence.

The recent publication of PEXIVAS randomised controlled trial triggered this guideline.[5] The Rapid Recommendations team felt that the results of this study, interpreted in the context of existing evidence, might change practice.

Our international guideline panel included patient partners with AAV with or without experience of plasma exchange, a caregiver for a patient who has ESKD, rheumatologists, nephrologists, an intensivist specialised in pulmonary vasculitis and pulmonary haemorrhage, a paediatrician specialised in vasculitis and autoinflammatory diseases, general internists, and methodologists (see appendix 4 on bmj.com for details of panel members). No panel member had relevant financial conflicts of interest; intellectual and professional conflicts were minimised and transparently described (see appendix 4 for details of competing interests).

The panel decided the scope of the recommendation and rated the outcome importance to patients. The panel judged the following as patient-important outcomes for decision making: mortality, ESKD, remission of AAV, health related quality of life, relapse of AAV, and serious infections (defined as infection requiring intravenous antibiotics or hospitalisation) and other serious adverse events associated with plasma exchange or glucocorticoids.

The panel met online to discuss the evidence and to formulate recommendations. The panel followed the *BMJ Rapid Recommendations* procedures for creating trustworthy guidelines[27], including using the GRADE approach to critically appraise the evidence and create recommendations (appendix 5 on bmj.com). The panel considered the balance of benefits, harms, and burdens and other practical issues related to plasma exchange and reduced dose regimen of glucocorticoids in the context of AAV, as well as typical and expected variations in patient values and preferences[28]. Within the GRADE approach, recommendations can be strong or weak (also known as conditional), for or against a course of action[29].

Box End

The recommendation is based on two linked systematic reviews on benefits and harms of plasma exchange and different dose regimens of glucocorticoids in patients with AAV[6,7]. The main infographic provides an overview of the relative and absolute benefits and harms of plasma exchange and reduced dose regimen of glucocorticoids in standard GRADE format. Box 3 shows articles linked to this guideline.

Box Start

Box 3. Linked resources for this *BMJ Rapid Recommendations* cluster

- Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ* 2022;376:e064597
  - Summary of the results from the Rapid Recommendation process
- Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: an updated systematic review and meta-analysis. *BMJ* 2022;376:e064604
  - Review and meta-analysis of randomised trials that assess effects of plasma exchange for antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV).
- Xiao Y, Guyatt G, Zeng L, et al. The comparative efficacy and safety of alternative glucocorticoids regimens in patients with ANCA-associated vasculitis: a systematic review. *BMJ Open* 2022; doi:10.1136/bmjopen-2021-050507
  - Review and meta-analysis of randomised trials that assess effects of alternative glucocorticoids regimens for AAV.
- Walsh M. Predicting the 1-year risk of kidney failure in ANCA associated vasculitis. BMJ Medicine [in review].
  - Prediction model of risk of kidney failure in AAV.
- MAGICapp (https://app.magicapp.org/#/guideline/4218)
  - Expanded version of the results with multilayered recommendations, evidence summaries, and decision aids for use on all devices.

Box End

Current practice

Who should use plasma exchange?

Most existing guidelines recommend in favour of plasma exchange in patients with severe kidney impairment (serum creatinine ≥500 µmol/L) or with active vasculitis despite ongoing remission induction therapy (see table 1)[8-15]. However, guidelines vary in the recommendation for patients with severe diffuse pulmonary haemorrhage, with some guidelines recommending in favour of plasma exchange, while others conclude there is insufficient evidence to support plasma exchange in these patients.

Table 1. Current recommendations for plasma exchange and dose regimen of glucocorticoids in patients with antineutrophil cytoplasmic antibody-associated vasculitis (AAV)

Organisation and year of publication	Recommendation of plasma exchange (PLEX) in		Recommendation of tapering regimen of glucocorticoids in induction therapy
	AAV and kidney involvement	AAV and pulmonary haemorrhage	
ASFA 2020[8]	*For patients with creatinine ≥500 µmol/L:* In favour of PLEX as accepted second line therapy alone or as adjuvant; support use	Consider PLEX for pulmonary haemorrhage a class I indication (accepted first line therapy) (strong	No recommendation

	of PLEX in select patients with biopsy proven RPGN (strong recommendation based on moderate quality evidence). *For patients with creatinine <500 µmol/L:* Optimal role not established, decision should be individualised (weak recommendation based on low or very low quality evidence)	recommendation based on low quality evidence)	
KDIGO 2020[9]	Against routine use of PLEX for patients with GFR <50 mL/min/1.73 m^2; PLEX can be considered for more severe presentations (serum creatinine >500 µmol/L, especially if oliguric)	In favour of PLEX for AAV and diffuse alveolar haemorrhage plus hypoxaemia	No explicit recommendation, but commented that (*a*) in most RCTs oral glucocorticoids started at 1 mg/kg/day; (*b*) PEXIVAS trial showed more rapid reduction was as effective but safer than "standard" corticosteroid tapering regimen
ARCH 2020[10]	In favour of PLEX for AAV and rapidly progressive glomerulonephritis	In favour of PLEX for AAV and pulmonary haemorrhage	No recommendation
Japan Research Committee of the Ministry of Health, Labour, and Welfare 2017[11]	In favour of PLEX for AAV and severe renal impairment	No recommendation	No recommendation
BSR 2017[12]	In favour of PLEX for AAV and rapidly progressive glomerulonephritis with serum creatinine >5.8 mg/dL	Insufficient evidence to support PLEX for AAV presenting with pulmonary haemorrhage, PLEX possibly beneficial	Prednisone or prednisolone prescribed at initial dose of 0.5-1.0 mg/kg/day (max 80 mg/day) for 1-4 weeks followed by tapering 10 mg every 2-4 weeks until 20 mg/day. Then taper dose 2.5-5.0 mg every 2-4 weeks until complete withdrawal
EULAR/ERA-EDTA 2016[13]	In favour of PLEX for AAV and serum creatinine level ≥500 mmol/L due to rapidly	In favour of PLEX for AAV and severe diffuse pulmonary haemorrhage	No recommendation

	progressive glomerulonephritis in new or relapsing disease		
CanVasc 2016[14]	Against PLEX as first line therapy for AAV and severe renal involvement (GFR <50 mL/min). PLEX may be a reasonable adjuvant therapy if patients clinically deteriorate	Against PLEX as first line therapy for AAV and pulmonary haemorrhage. PLEX may be a reasonable adjuvant therapy if patients clinically deteriorate	No recommendation
BSR/BHPR 2014[15]	In favour of PLEX for AAV and severe renal failure (serum creatinine >500 mmol/L)	In favour of PLEX for AAV and pulmonary haemorrhage	Glucocorticoids usually given as daily oral prednisolone, initially at high doses (1 mg/kg up to 60 mg) with dose rapidly reduced to 15 mg prednisolone at 12 weeks

ASFA = American Society for Apheresis; Kidney Disease: KDIGO = Improving Global Outcomes; ARCH = Arthritis Research and Collaboration Hub; BSR = Brazilian Society of Rheumatology; EULAR/ERA-EDTA = European League Against Rheumatism/European Renal Association-European Dialysis and Transplant Association; CanVasc = Canadian Vasculitis Research Network; BSR/BHPR = British Society for Rheumatology/British Health Professionals in Rheumatology; GFR = glomerular filtration rate, RCT = randomised controlled trial.

What is the tapering regimen of glucocorticoids for the first six months of therapy?

In guidelines that have a recommendation on the dose regimen of glucocorticoids, the initial dose of prednisolone or equivalent is 0.5-1 mg/kg/day. There is, however, no standard for the taper rate of glucocorticoids after initial treatment. A guideline from the British Society for Rheumatology/British Health Professionals in Rheumatology recommends a “rapid reduction” of glucocorticoids after the initial dose[15]. The recommended taper rate is, however, slower than the reduced dose regimen in the PEXIVAS trial.

A review of the prednisolone dose regimen in trials compared the dose in the PEXIVAS trial with those in other key trials. On average, a dose of 10 mg was achieved after 19 weeks in the standard dose regimen group of the PEXIVAS trial and in other trials, and after 13 weeks in the reduced dose regimen group of PEXIVAS. The standard dose regimen achieved a dose of 7.5 mg after 21 weeks, while the reduced dose regimen achieved this dose four weeks earlier (after 17 weeks) (see appendix 1 for more details). A cross sectional survey among 34 hospitals in England revealed a large variation in the initial dose and taper rate of glucocorticoids in patients with AAV.[16]

The evidence

What are the benefits and harms of plasma exchange in patients with AAV, with or without kidney involvement?

We incorporated the PEXIVAS trial into a linked systematic review to generate pooled estimates of effect (see infographic). The review included nine RCTs and 1060 patients with AAV

comparing plasma exchange in addition to standard care (that is, immunosuppression and glucocorticoids) versus standard care alone. Table 2 provides an overview of the trials and participants. PEXIVAS, the largest of the nine trials, evaluated the effect of plasma exchange in 704 patients with severe AAV. The systematic review analysed mortality and ESKD separately, rather than as composite.

Table 2. Characteristics of 9 randomised controlled trials (1060 patients) included in systematic review of plasma exchange in patients with antineutrophil cytoplasmic antibody-associated vasculitis (AAV), with or without kidney involvement

	Values
Trial characteristics	**Mean (range) of means across trials**
No of patients enrolled	118 (14-704)
Length of follow-up (months)	Median 36 (12-127)
Dose regimen of plasma exchange	Centrifugation or filter separation; 8 trials used albumin and/or crystalloid replacement solution for a median 8 treatments; exchange volume ranged from 1 to 1.5 plasma volumes (or 40 to 60 mL/kg or fixed volume of 3.5-4 L)
Setting	Multiple centres internationally including Europe, North America, and Australasia
Funding	Public funding only (4 trials) In-kind supplies from industry partner (1 trial) Public funding and in-kind supplies from three industry partners (1 trial) Not reported (3 trials)
Patient involvement	No trial reported patient involvement in design or conduct
Patient characteristics	**Mean (range) of means across trials**
Age (years)	56 (47-67)
Sex (% women)	35 (22-44)
ANCA positive	84% in 6 trials that measured ANCA
Kidney function (serum creatinine concentration µmol/L)	Median 716 (256-1176)
Presence of pulmonary haemorrhage	Patients with pulmonary haemorrhage included (4 trials) Patients with severe pulmonary haemorrhage included (1 trial)

We used the control group event rates in the systematic review to estimate the baseline risks for outcomes of mortality, serious infections, and relapse of AAV, and used the data from seven multinational RCTs conducted by the European Vasculitis Study Group with 798 patients to estimate the baseline risk for the outcome of ESKD[17-23]. The systematic review found no credible evidence that the relative effect of plasma exchange would vary on the basis of kidney function or pulmonary haemorrhage[6]. We therefore used the baseline risks, along with the pooled relative risk for overall patients at the timeline of one year and long term follow-up (median 3 years) from the systematic review, to calculate the absolute effect estimates presented in our evidence summaries.

Mortality and relapse of AAV

Plasma exchange probably has little or no effect on mortality (risk difference (RD) 0.8% reduction (95% CI 3.9% reduction to 3.6% increase) at 1 year; RD 1.3% reduction (5.5% reduction to 3.6% increase) at long term follow-up; both moderate certainty due to imprecision). Plasma exchange may reduce relapse of AAV (RD 2.1% reduction (11.6% reduction to 13.9% increase) at long term follow-up; low certainty due to very serious imprecision; no data available at 1 year).

End stage kidney disease and serious infections

The absolute effects of plasma exchange in ESKD and serious infections vary significantly with baseline risks. The panel, therefore, decided to use risk of developing ESKD at 1 year (i.e. baseline risk) as a stratification variable to present the absolute effects of plasma exchange on ESKD and serious infections, and then discussed the tradeoff between benefits and harms in each of the risk groups.

The panel stratified the risks of developing ESKD for four groups (see table 3). A linked prognostic study demonstrates that serum creatinine, as a single predictor, can provide robust estimates of the risk of developing ESKD[24]. Patients with serum creatinine ≤200 µmol/L, >200-300 µmol/L, >300-500 µmol/L and >500 µmol/L fall, respectively, into low, low-moderate, moderate-high, and high risk groups (see infographic)[24]. The panel recognised that, although serum creatinine level could well predict the risk of ESKD[24], using serum creatinine as a single predictor has limitations (for example, the serum creatinine might change rapidly or the prognosis may be modified by tests such as kidney biopsy).

Because of availability of baseline risk strata, the linked systematic review provided the absolute effects of plasma exchange in ESKD and serious infections in a time frame of 1 year rather than a longer time frame[6]. Plasma exchange probably reduces the 1 year risk of ESKD (the absolute risk reduction approximately 0.1% in low risk group, 2.1% reduction in low-moderate risk group, 4.6% reduction in moderate-high risk group, and 16.0% in high risk group or patients requiring dialysis) but increases the risk of serious infections (the absolute risk increase approximately 2.7% in low risk group, 4.9% in low-moderate risk group, 8.5% increase in moderate-high risk group, and 13.5% in high risk group or patients requiring dialysis) at 1 year (moderate to high certainty). See infographic for more details.

What are the benefits and harms of plasma exchange in patients with AAV and pulmonary haemorrhage without kidney involvement?

In patients with pulmonary haemorrhage without kidney involvement, the key benefit outcome becomes risk reduction in death, and the key harm outcome remains an increase in

serious infections. Because we have limited data regarding the baseline risks of death and serious infections in this group of patients, we estimated the baseline risk for outcome of mortality in a time frame of 1 year using the average mortality (20.8%) in patients with pulmonary haemorrhage in the control group of the PEXIVAS trial. The estimate comes from a mix of patients with or without kidney involvement. Thus, this mortality (20.8%) might overestimate mortality for the average patient with pulmonary haemorrhage without kidney involvement. We estimated the baseline risk of serious infections as similar to the risk in the entire control group of the RCTs (25%).

We are uncertain whether plasma exchange has an effect on death at 1 year (RD 1.5% reduction (95% CI 7.1% reduction to 6.4% increase); very low certainty due to indirectness and very serious imprecision). It may have an important increase in serious infections at 1 year (RD 6.8% increase (95% CI 0.8% increase to 14% increase); low certainty due to indirectness and imprecision).

What are the benefits and harms of reduced dose regimen of glucocorticoids?

The linked systematic review of comparative efficacy and safety of alternative glucocorticoids regimen included two RCTs at low risk of bias. One trial included 704 patients with severe AAV; the other included 140 patients with newly diagnosed AAV (of which 134 patients completed the trial)[22, 32]. Due to the heterogeneity in the population and in the regimens of glucocorticoids, the systematic review authors descriptively presented the two trials and did not combine the results using meta-analysis.

Compared with standard dose regimen of oral glucocorticoids, the reduced dose regimen of oral glucocorticoids probably has an important reduction in serious infections at a follow-up of 6 months to 1 year (RD 5.9% to 12.8% reduction; moderate certainty due to imprecision), and may reduce death from any cause at long term follow-up (RD 1.7% to 2.1% reduction; low certainty due to very serious imprecision) without increasing the risk of ESKD (RD 1.5% reduction to 0.4% increase; moderate certainty due to imprecision). The reduced dose regimen probably has little or no effect on disease remission, relapse, or health related quality of life (moderate to high certainty).

Values and preferences

To elicit the guideline panel's view of patients' values and preferences (primarily the relative value patients would place on avoiding ESKD and avoiding serious infections) we conducted two formal panel surveys. In the first survey, conducted before the panel reviewed the evidence of benefits (that is, reduction in ESKD), the panel members (including four patients and one care giver partner), presented with the harms associated with plasma exchange, expressed their view of the magnitude of reduction in ESKD that patients would require to choose plasma exchange (see appendix 2 for details of the survey process and results). In that survey and the subsequent discussion, the panel concluded that patients would place a high value on reduction in ESKD, and less value on avoiding serious infections.

For making a judgment about how patients with varying risks of developing ESKD would view the trade-off between benefit (that is, reduction in ESKD) and harm (increase in serious infections) of plasma exchange, the panel completed a second survey. In this survey, they considered the estimated absolute effects of plasma exchange on the key benefit and the key

harm from the linked systematic review (see appendix 3 for details of the survey process and results). Based on the survey and panel discussion, the panel agreed that, for patients with low or low-moderate risk of developing ESKD, the harms of serious infections outweighed the benefits in terms of reduction in ESKD; but, because it was a close balance, the majority of patients but not all (50-90%) would decline plasma exchange. The panel agreed that, for patients with moderate-high or high risk of developing ESKD or requiring dialysis, the benefits outweigh the harms, such that the majority of patients would choose plasma exchange.

Understanding the recommendations

Recommendation 1. We suggest immunosuppression alone rather than adding plasma exchange for patients with AAV and low or low-moderate risk of developing ESKD, with or without pulmonary haemorrhage (weak recommendation)

This recommendation applies to adult patients with AAV and with low or low-moderate risk of ESKD with or without pulmonary haemorrhage. Following GRADE guidance, a weak recommendation implies that the majority (50-90%) of patients would decline plasma exchange, but a minority (<50%) would, depending on individual shared decision making, choose to receive plasma exchange.

The panel made this recommendation on the basis that, for the majority of patients, moderate to high certainty evidence of a reduction in ESKD (0.1% to 2.1% reduction) in patients with low or low-moderate risk of ESKD does not counterbalance the increase in serious infections (2.7% to 4.9% increase) over a timeframe of 1 year.

Recommendation 2. We suggest plasma exchange plus immunosuppression rather than immunosuppression alone for patients with AAV and moderate-high or high risk of developing ESKD or requiring dialysis, with or without pulmonary haemorrhage (weak recommendation)

This recommendation applies to adult patients with AAV and with moderate-high or high risk of ESKD or requiring dialysis with or without pulmonary haemorrhage. A weak recommendation implies that most patients (50-90%) would choose plasma exchange; a minority (<50%) would, depending on individual shared decision making, decline plasma exchange.

The panel made this recommendation on the basis of moderate to high certainty evidence of an important reduction in ESKD (4.6% to 16.0% reduction) and an important increase in serious infections (8.5% to 13.5% increase) in patients with moderate-high to high risk of ESKD or requiring dialysis. The panel considered patients would generally place more value on avoiding ESKD and less value on avoiding risk of serious infections.

Recommendation 3. We suggest immunosuppression alone without plasma exchange in patients with AAV and pulmonary haemorrhage without kidney involvement (weak recommendation)

This recommendation applies to adult patients with AAV and pulmonary haemorrhage without kidney involvement, and does not apply to those with kidney involvement. For the latter, please refer to recommendations 1 and 2 in this guideline.

A weak recommendation for immunosuppression alone reflects the panel's view that the majority (50-90%) of patients with AAV and isolated pulmonary haemorrhage without kidney

involvement would decline plasma exchange; a minority (<50%) of patients would, depending on individual shared decision making, choose plasma exchange.

The panel made this recommendation based on indirect evidence that plasma exchange may increase the risk of serious infections (6.8% increase) but uncertainty about the effect on death (1.5% reduction with very wide confidence interval) over a timeframe of 1 year.

Recommendation 4. We recommend reduced dose regimen of glucocorticoids rather than standard dose regimen of glucocorticoids during the first six months of therapy in patients with AAV (strong recommendation)

The panel recognised that the evidence that supports the reduced dose regimen of glucocorticoids is based on the systematic review of reduced dose versus standard dose of glucocorticoids in patients with severe AAV and patients with newly diagnosed AAV[7].

The panel made this recommendation on a basis of moderate certainty evidence of an important reduction in serious infections (5.9% to 12.8% reduction) and no increase in death or ESKD (2.1% reduction for death and 0.4% increase for ESKD) in patients with severe AAV over a timeframe of 1 year, and similar findings in patients with newly diagnosed AAV. The panel considered the strong recommendation mandated by the decreased harm and no decreased benefit. Standard dose regimen of glucocorticoids may be appropriate for patients who do not respond to a reduced dose regimen.

Practical issues

Tables 4 and 5 outline the key practical issues regarding the use of plasma exchange and reduced dose regimen of glucocorticoids in patients with AAV. The protocols for either plasma exchange or dose regimen of glucocorticoids might vary largely between medical institutions. Patients using plasma exchange need intravenous lines or central venous catheters that may cause discomfort or increase the risk of infection, clotting, or bleeding, and might need blood transfusions.

Table 3. Practical issues regarding use of plasma exchange in patients with antineutrophil cytoplasmic antibody-associated vasculitis

Practical issues	Plasma exchange + standard care	Standard care
Procedure and device	Heterogeneity in plasma exchange protocols	Null
Coordination of care	Need for an intravenous line with plasma exchange, which may cause discomfort, infection, or bleeding	Null
Coordination of care	Potential need for blood products with plasma exchange.	Null
Adverse effects, interactions, and antidote	Plasma exchange may affect the pharmacokinetics of some drugs	Null
Costs and access	Potential need for transfer to another centre to get plasma exchange Cost of plasma exchange is high and might not be covered by medical insurance	Null

Table 4. Practical issues regarding use of reduced dose regimen of oral glucocorticoids (prednisone or prednisolone) in patients with antineutrophil cytoplasmic antibody-associated vasculitis

Practical issues	Reduced dose regimen	Standard dose regimen
Medication routine	• Initial dose in 1st week is same as that in standard dose regimen • In 2nd week, dose is reduced by ~50% • In 3rd to 6th weeks, dose is reduced by 5 mg in every 2 weeks • In 7th to 14th weeks, dose is reduced by 2.5-1 mg every 2 weeks until reaches 5 mg/day at 15th week • At 6 months, cumulative dose of oral glucocorticoids is <60% of that in standard dose regimen group	• Initial dose in first 2 weeks is: Patients <50 kg weight, 50 mg/day Patients 50-75 kg: 60 mg/day Patients >75 kg, 75 mg/day • From 3rd to 6th week, dose reduced by 10 mg every 2 weeks • From 7th to 22nd week, dose is reduced by 5-2.5 mg every 2-4 weeks until reaches 5 mg/day at 23rd week
Medication routine	Patients intolerant of oral glucocorticoids or for whom oral glucocorticoids are contraindicated could be given an equivalent daily intravenous dose	
Adverse effects, interactions, and antidote	Adverse events of glucocorticoids including impaired fasting glucose, loss of bone mineral density, fractures, weight gain, mood changes, etc	

Cost and resources

In some jurisdictions the cost of plasma exchange might not be covered by medical insurance, and access might be limited.

Uncertainty

- The process of determining the threshold at which the recommendation changes from immunosuppression alone to adding plasma exchange proved challenging.
- The uncertainty in the estimates of risk of ESKD: although the linked prognostic study showed that serum creatinine as a single predictor can effectively predict the risk of ESKD in patients with AAV,[24] a prognostic model with multiple and more stable predictors is likely to improve prediction and thus risk stratification.
- The uncertainty in patients' values and preferences regarding the trade-off between benefit (reduction in ESKD) and harm (increase in serious infections). A broader patient survey would be helpful in ascertaining patients' values and preferences.
- The extent to which the safety and efficacy of the recommended regimen, which included intravenous glucocorticoids before beginning the reduced dose regimen, to regimens that do not include intravenous glucocorticoids is uncertain
- Very limited data proved available to estimate risk of death or serious infections in patients with AAV and pulmonary haemorrhage without kidney involvement.

- The benefits and harms of plasma exchange and reduced dose regimen of glucocorticoids in patients with both antineutrophil cytoplasmic and anti-glomerular basement membrane antibodies was not evaluated in this review, and these recommendations do not apply to them.
- Other than infections, serious adverse events associated with plasma exchange (such as allergic reactions, cardiovascular events) remain uncertain. As the rate of these serious adverse events is low, current RCTs are under-powered to detect differences.
- The dose regimens of glucocorticoids vary in clinical practice. The comparative efficacy and safety of regimens other than those tested remain uncertain.

Update to this article

Table 6 shows evidence that has emerged since the publication of this article. A group will assess new evidence as it becomes available and make a judgment as to whether it might alter recommendations.

Table 5. New evidence which has emerged after initial publication

Date	New evidence	Citation	Findings	Implications for recommendation(s)
There are currently no updates to the article				

Box Start

How patients were involved in the creation of this article

Four patient partners with ANCA-associated vasculitis with or without experience of plasma exchange and a caregiver for a patient who has end stage kidney disease were full panel members. These panel members identified important outcomes, participated in the teleconferences and email discussions on the evidence and recommendation. They also contributed to the identification of practical issues related to the decision of plasma exchange and glucocorticoids regimen and met all authorship criteria for the present article. We thank them for their time and contribution.

Box End

Funding: This guideline was not funded.

Competing interests: All authors have completed the *BMJ Rapid Recommendations* interest disclosure form and a detailed, contextualised description of all disclosures is reported in appendix 4. As with all *BMJ Rapid Recommendations*, the executive team and *The BMJ* judged that no panel member had any financial conflict of interest. Professional and academic interests were minimised as much as possible, while maintaining necessary expertise on the panel to make fully informed decisions.

Participation in the panel and authorship of this manuscript does not constitute organisational endorsement of the recommendations.

Transparency: L Zeng and A Mahr affirm that the manuscript is an honest, accurate, and transparent account of the recommendation being reported; that no important aspects of the

recommendation have been omitted; and that any discrepancies from the recommendation as planned have been explained.

Licence for publication: The corresponding author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the *BMJ* Publishing Group Ltd to permit this article (if accepted) to be published in *BMJ* and any other *BMJPGL* products and sublicences such use and exploit all subsidiary rights, as set out in our licence (http://group.bmj.com/products/journals/instructions-for-authors/licence-forms).

References

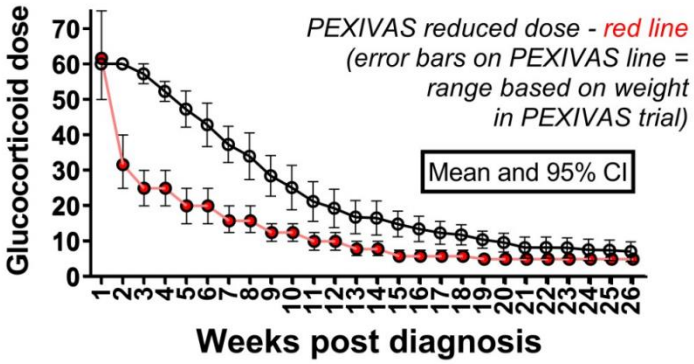
1. Lazarus B, John GT, O'Callaghan C, et al. Recent advances in anti-neutrophil cytoplasmic antibody-associated vasculitis. *Indian J Nephrol.* 2016;26:86-96. doi:10.4103/0971-4065.171225
2. Rhee RL, Hogan SL, Poulton CJ, et al. Trends in long-term outcomes among patients with antineutrophil cytoplasmic antibody-associated vasculitis with renal disease. *Arthritis Rheumatol.* 2016;68:1711-20. doi:10.1002/art.39614
3. Flossmann O, Berden A, de Groot K, et al; European Vasculitis Study Group. Long-term patient survival in ANCA-associated vasculitis. *Ann Rheum Dis.* 2011;70:488-94. doi:10.1136/ard.2010.137778
4. Lai QY, Ma TT, Li ZY, et al. Predictors for mortality in patients with antineutrophil cytoplasmic autoantibody-associated vasculitis: a study of 398 Chinese patients. *J Rheumatol.* 2014;41:1849-55. doi:10.3899/jrheum.131426
5. Walsh M, Merkel PA, Peh CA, et al; PEXIVAS Investigators. Plasma exchange and glucocorticoids in severe ANCA-associated vasculitis. *N Engl J Med.* 2020;382:622-31. doi:10.1056/NEJMoa1803537
6. Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: An updated systematic review and meta-analysis. *BMJ.* 2022;376:e064604. doi:10.5414/CN110410
7. Xiao Y, Guyatt G, Zeng L, et al. The comparative efficacy and safety of alternative glucocorticoids regimens in patients with ANCA-associated vasculitis: a systematic review. *BMJ Open.* 2022;12(2):e050507. doi:10.1136/bmjopen-2021-050507

8. Balogun RA, Sanchez AP, Klingel R, et al. Update to the ASFA guidelines on the use of therapeutic apheresis in ANCA-associated vasculitis. *J Clin Apher.* 2020;35:493-9. PubMed doi:10.1002/jca.21820
9. KDIGO. KDIGO clinical practice guideline on glomerular disease. 2020. https://kdigo.org/wp-content/uploads/2017/02/KDIGO-GN-GL-Public-Review-Draft_1-June-2020.pdf.
10. Dirikgil E, Tas SW, Rutgers A, et al. A Dutch consensus statement on the diagnosis and treatment of ANCA-associated vasculitis. *Neth J Med.* 2020;78:71-82.
11. Harigai M, Nagasaka K, Amano K, et al. 2017 clinical practice guidelines of the Japan Research Committee of the Ministry of Health, Labour, and Welfare for Intractable Vasculitis for the management of ANCA-associated vasculitis. *Mod Rheumatol.* 2019;29:20-30. doi:10.1080/14397595.2018.1500437
12. Souza AWS, Calich AL, Mariz HA, et al. Recommendations of the Brazilian Society of Rheumatology for the induction therapy of ANCA-associated vasculitis. *Rev Bras Reumatol Engl Ed.* 2017;57(Suppl 2):484-96. doi:10.1016/j.rbre.2017.06.003
13. Yates M, Watts RA, Bajema IM, et al. EULAR/ERA-EDTA recommendations for the management of ANCA-associated vasculitis. *Ann Rheum Dis.* 2016;75:1583-94. doi:10.1136/annrheumdis-2016-209133
14. McGeoch L, Twilt M, Famorca L, et al; Canadian Vasculitis Research Network. CanVasc recommendations for the management of antineutrophil cytoplasm antibody-associated vasculitides. *J Rheumatol.* 2016;43:97-120. doi:10.3899/jrheum.150376
15. Ntatsaki E, Carruthers D, Chakravarty K, et al; BSR and BHPR Standards, Guidelines and Audit Working Group. BSR and BHPR guideline for the management of adults with ANCA-associated vasculitis. *Rheumatology (Oxford).* 2014;53:2306-9. doi:10.1093/rheumatology/ket445
16. Pearce FA, McGrath C, Sandhu R, et al. Outcomes and compliance with standards of care in anti-neutrophil cytoplasmic antibody-associated vasculitis-insights from a large multiregion audit. *Rheumatol Adv Pract.* 2018;2:rky025. doi:10.1093/rap/rky025
17. Pepper RJ, Chanouzas D, Tarzi R, et al; European Vasculitis Study (EUVAS) investigators. Intravenous cyclophosphamide and plasmapheresis in dialysis-dependent ANCA-associated vasculitis. *Clin J Am Soc Nephrol.* 2013;8:219-24. PubMed doi:10.2215/CJN.03680412

18. Jayne DR, Gaskin G, Rasmussen N, et al; European Vasculitis Study Group. Randomized trial of plasma exchange or high-dosage methylprednisolone as adjunctive therapy for severe renal vasculitis. *J Am Soc Nephrol.* 2007;18:2180-8. PubMed doi:10.1681/ASN.2007010090
19. Jayne D, Rasmussen N, Andrassy K, et al; European Vasculitis Study Group. A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. *N Engl J Med.* 2003;349:36-44. doi:10.1056/NEJMoa020286
20. De Groot K, Rasmussen N, Bacon PA, et al. Randomized trial of cyclophosphamide versus methotrexate for induction of remission in early systemic antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum.* 2005;52:2461-9. doi:10.1002/art.21142
21. Hiemstra TF, Walsh M, Mahr A, et al; European Vasculitis Study Group (EUVAS). Mycophenolate mofetil vs azathioprine for remission maintenance in antineutrophil cytoplasmic antibody-associated vasculitis: a randomized controlled trial. *JAMA.* 2010;304:2381-8. doi:10.1001/jama.2010.1658
22. Jones RB, Tervaert JW, Hauser T, et al; European Vasculitis Study Group. Rituximab versus cyclophosphamide in ANCA-associated renal vasculitis. *N Engl J Med.* 2010;363:211-20. doi:10.1056/NEJMoa0909169
23. Jones RB, Hiemstra TF, Ballarin J, et al; European Vasculitis Study Group (EUVAS). Mycophenolate mofetil versus cyclophosphamide for remission induction in ANCA-associated vasculitis: a randomised, non-inferiority trial. *Ann Rheum Dis.* 2019;78:399-405. doi:10.1136/annrheumdis-2018-214245
24. Walsh M. POS-167: predicting the 1-year risk of kidney failure in ANCA associated vasculitis. *Kidney Int Rep.* 2021;6:S68.
25. Geetha D, Jefferson JA. ANCA-associated vasculitis: core curriculum 2020. *Am J Kidney Dis.* 2020;75:124-37. doi:10.1053/j.ajkd.2019.04.031
26. Yates M, Watts R. ANCA-associated vasculitis. *Clin Med (Lond).* 2017;17:60-4. doi:10.7861/clinmedicine.17-1-60
27. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. *BMJ.* 2016;354:i5191. doi:10.1136/bmj.i5191

28. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013;66:726-35. doi:10.1016/j.jclinepi.2013.02.003
29. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. Going from evidence to recommendations. *BMJ.* 2008;336:1049-51. doi:10.1136/bmj.39493.646875.AE</jrn>
30. Carpenter S, Tervaert JWC, Yacyshyn E, et al. Advances in therapeutic treatment options for ANCA-associated vasculitis. *Expert Opin Orphan Drugs.* 2020;8:127-36 doi:10.1080/21678707.2020.1760837.

Appendix 1 Glucocorticoid (prednisolone or equivalent) dose regimen in key trials: PEXIVAS versus other 9 trials



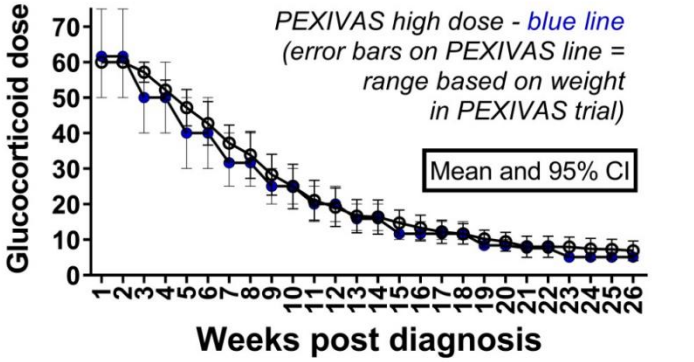


Figure legend: The other 9 trials include:

1. Jayne D, Rasmussen N, Andrassy K, et al. A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. *N Engl J Med.* 2003;349:36–44.

2. De Groot K, Rasmussen N, Bacon PA, et al. Randomized trial of cyclophosphamide versus methotrexate for induction of remission in early systemic antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum.* 2005;52:2461–9.

3. Stassen PM, Tervaert JWC, Stegeman CA. Induction of remission in active anti-neutrophil cytoplasmic antibody-associated vasculitis with mycophenolate mofetil in patients who cannot be treated with cyclophosphamide. *Ann Rheum Dis.* 2007;66:798–802.

4. Cohen P, Pagnoux C, Mahr A, et al. Churg-Strauss syndrome with poor-prognosis factors: a prospective multicenter trial comparing glucocorticoids and six or twelve cyclophosphamide pulses in forty-eight patients. *Arthritis Rheum.* 2007;57:686–93.

5. ayne DRW, Gaskin G, Rasmussen N, et al. Randomized trial of plasma exchange or high-dosage methylprednisolone as adjunctive therapy for severe renal vasculitis. *J Am Soc Nephrol.* 2007;18:2180–8.

6. de Groot K, Harper L, Jayne DR, et al. Pulse versus daily oral cyclophosphamide for induction of remission in antineutrophil cytoplasmic antibody-associated vasculitis: a randomized trial. *Ann Intern Med.* 2009;150:670–80.

7. Jones RB, Tervaert JW, Hauser T, et al. Rituximab versus cyclophosphamide in ANCA-associated renal vasculitis. *N Engl J Med.* 2010;363:211–20.
8. Stone JH, Merkel PA, Spiera R, et al. Rituximab versus cyclophosphamide for ANCA-associated vasculitis. *N Engl J Med.* 2010;363:221–32.
9. Mansfield N, Hamour S, Habib AM, et al. Prolonged disease-free remission following rituximab and low-dose cyclophosphamide therapy for renal ANCA-associated vasculitis. *Nephrol Dial Transplant.* 2011;26:3280–6.

Appendix 2 First panel survey of values and preferences towards plasma exchange in patients with ANCA-associated vasculitis (AAV)

Before the first panel meeting, 21 panel members completed an online survey asking them to, based on their experience, hypothesize the magnitude of reduction in end-stage kidney disease that, given its associated increase in serious infections, patients would require to choose plasma exchange. To minimize the influence of preconceived beliefs regarding the advisability of using plasma exchange, we did not present, prior to or in the survey, the estimated effect of plasma exchange in reduction of end-stage kidney disease.

One patient and one caregiver completed the questionnaire together, so we received 20 responses in total. The panel members stopped answering questions when they moved from "majority would decline plasma exchange" to "majority would choose plasma exchange" (i.e. they identified the threshold for benefit). Therefore, the total number of votes for individual questions sometimes less than 20. The following presents the survey questionnaire followed by panel members' votes. Finally, we summarize the thresholds identified by individual panel members.

Introduction

Plasma exchange may reduce the risk of some outcomes, such as end-stage kidney disease but at the cost of an increased risk of adverse events. We need to understand what degree of benefit (in terms of a reduced risk of end-stage kidney disease) would be required to accept the risks of adverse events. Below we will provide information on the potential risks of plasma exchange. You will then be presented with a series of questions asking if you believe patients would choose to receive plasma exchange under a specific scenario in terms of reducing end-stage kidney disease but increasing in the risk of serious infection. Each question will vary the risk reduction of end-stage kidney disease and the risk increase of serious infection. Please read these carefully.

"Values and preferences regarding plasma exchange for adults with ANCA-associated vasculitis and kidney involvement and/or diffuse pulmonary hemorrhage differ markedly. Some patients might choose to receive plasma exchange with a very small benefit on the absolute risk reduction of end-stage kidney disease, and some will be reluctant to take it even if there is a large benefit."

Would you agree or disagree with this statement?*

- Agree: 16
- Disagree: 4

Here we give a summary of the harms, burdens and benefits other than risk reduction on end-stage kidney disease of plasma exchange that you should bear in mind when you answer the follow questions.

<u>Harms and burdens:</u>

- Plasma exchange increased the risk of serious infection (defined as requiring hospitalization or intravenous antibiotics) compared with usual care in the first year. The absolute risk increase is 6 more per 100 patients.
- Plasma exchange may result in the transmission of serious viral infections such as hepatitis B, hepatitis C or HIV in less than 1 in 7 million patients.

- Plasma exchange results in transfusion related acute lung injury in about 1 in 12,000 patients.
- Plasma exchange is performed several times during a period of approximate 2 weeks.
- Some patients receiving plasma exchange need central line placement.

Benefits other than risk reduction on end-stage kidney disease:

- Plasma exchange did not affect death or disease relapse at 12 months or after longer follow-up.

In the first year, a patient who takes plasma exchange has a **1 in 100 (1%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 3
- Majority (51-74%) would **choose** plasma exchange: 2
- Majority (51-74%) would **decline** plasma exchange: 5
- Most (75-90%) would **decline** plasma exchange: 4
- All or almost all (over 90%) would **decline** plasma exchange: 6

In the first year, a patient who takes plasma exchange, has a **15 in 100 (15%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 6
- Most (75- 90%) would **choose** plasma exchange: 5
- Majority (51-74%) would **choose** plasma exchange: 2
- Majority (51-74%) would **decline** plasma exchange: 2
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **2 in 100 (2%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 0
- Majority (51-74%) would **choose** plasma exchange: 0
- Majority (51-74%) would **decline** plasma exchange: 4
- Most (75-90%) would **decline** plasma exchange: 6
- All or almost all (over 90%) would **decline** plasma exchange: 3

In the first year, a patient who takes plasma exchange, has a **14 in 100 (14%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other

risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 6
- Most (75- 90%) would **choose** plasma exchange: 3
- Majority (51-74%) would **choose** plasma exchange: 4
- Majority (51-74%) would **decline** plasma exchange: 0
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **3 in 100 (3%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 1
- Majority (51-74%) would **choose** plasma exchange: 2
- Majority (51-74%) would **decline** plasma exchange: 3
- Most (75-90%) would **decline** plasma exchange: 7
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **13 in 100 (13%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 5
- Most (75- 90%) would **choose** plasma exchange: 2
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 0
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **4 in 100 (4%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 0
- Majority (51-74%) would **choose** plasma exchange: 2
- Majority (51-74%) would **decline** plasma exchange: 4
- Most (75-90%) would **decline** plasma exchange: 4
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **12 in 100 (12%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other

risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 2
- Most (75- 90%) would **choose** plasma exchange: 3
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 0
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **5 in 100 (5%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 1
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 2
- Most (75-90%) would **decline** plasma exchange: 2
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **11 in 100 (11%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 1
- Majority (51-74%) would **choose** plasma exchange: 4
- Majority (51-74%) would **decline** plasma exchange: 0
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **6 in 100 (6%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 0
- Majority (51-74%) would **choose** plasma exchange: 0
- Majority (51-74%) would **decline** plasma exchange: 4
- Most (75-90%) would **decline** plasma exchange: 1
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **10 in 100 (10%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other

risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 1
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 1
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **7 in 100 (7%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 0
- Majority (51-74%) would **choose** plasma exchange: 1
- Majority (51-74%) would **decline** plasma exchange: 3
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **9 in 100 (9%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 0
- Majority (51-74%) would **choose** plasma exchange: 0
- Majority (51-74%) would **decline** plasma exchange: 3
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **8 in 100 (8%) lower risk of end-stage kidney disease** and a **6 in 100 (6%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 0
- The majority would **decline** plasma exchange: 0
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

Now consider that the risk of serious infection is 18% instead of 6% and other burdens and harms were the same. Now, what would patients choose.

In the first year, a patient who takes plasma exchange, has a **1 in 100 (1%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 3
- The majority would **decline** plasma exchange: 3
- Most would **decline** plasma exchange: 4
- All or almost all would **decline** plasma exchange: 10

In the first year, a patient who takes plasma exchange, has a **15 in 100 (15%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 4
- Most would **choose** plasma exchange: 5
- The majority would **choose** plasma exchange: 3
- The majority would **decline** plasma exchange: 3
- Most would **decline** plasma exchange: 2
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **2 in 100 (2%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 1
- The majority would **decline** plasma exchange: 3
- Most would **decline** plasma exchange: 3
- All or almost all would **decline** plasma exchange: 5

In the first year, a patient who takes plasma exchange, has a **14 in 100 (14%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 2
- Most would **choose** plasma exchange: 6
- The majority would **choose** plasma exchange: 3
- The majority would **decline** plasma exchange: 0
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **3 in 100 (3%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 1
- The majority would **decline** plasma exchange: 3
- Most would **decline** plasma exchange: 4
- All or almost all would **decline** plasma exchange: 3

In the first year, a patient who takes plasma exchange, has a **13 in 100 (13%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 2
- Most would **choose** plasma exchange: 6
- The majority would **choose** plasma exchange: 1
- The majority would **decline** plasma exchange: 1
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **4 in 100 (4%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 0
- The majority would **decline** plasma exchange: 4
- Most would **decline** plasma exchange: 4
- All or almost all would **decline** plasma exchange: 1

In the first year, a patient who takes plasma exchange, has a **12 in 100 (12%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 7
- The majority would **choose** plasma exchange: 2
- The majority would **decline** plasma exchange: 0
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **5 in 100 (5%) lower risk of end-stage kidney disease** and **an 18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 1
- The majority would **decline** plasma exchange: 5
- Most would **decline** plasma exchange: 3
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **11 in 100 (11%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 5
- The majority would **choose** plasma exchange: 3
- The majority would **decline** plasma exchange: 0
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **6 in 100 (6%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 2
- The majority would **decline** plasma exchange: 6
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **10 in 100 (10%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 4
- The majority would **choose** plasma exchange: 2
- The majority would **decline** plasma exchange: 0
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **7 in 100 (7%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 3
- The majority would **decline** plasma exchange: 3
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **9 in 100 (9%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection** in addition to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 2
- The majority would **decline** plasma exchange: 1
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange: 0

In the first year, a patient who takes plasma exchange, has a **8 in 100 (8%) lower risk of end-stage kidney disease** and an **18 in 100 (18%) increased risk of serious infection in addition** to the other risks noted above. Given the harms and burdens of plasma exchange, how would patients view such benefits?

- All or almost all would **choose** plasma exchange: 0
- Most would **choose** plasma exchange: 0
- The majority would **choose** plasma exchange: 1
- The majority would **decline** plasma exchange: 1
- Most would **decline** plasma exchange: 0
- All or almost all would **decline** plasma exchange:0

Summary: **Given two levels of harm (increase in serious infections), the threshold for benefit (reduction in end-stage kidney disease) identified by individual panel members**

Panel member number	The threshold for reduction in end-stage kidney disease (expressed as the number reduced end-stage kidney disease per 100 patients)	
	6% increased risk of serious infections	18% increased risk of serious infections
#1	1	9
#2	1	1

#3	1	1
#4	1	2
#5	1	1
#6	3	15
#7	3	7
#8	3	3
#9	4	7
#10	4	5
#11	5	7
#12	5	8
#13	5	14
#14	7	10
#15	10	15
#16	10	15
#17	10	15
#18	11	6
#19	15	6
#20	15	15
Range	1-15	1-15
Median	4.5	7

Conclusions

Based on the results of the first survey, the panel acknowledged a likely large variation in patients' values and preferences regarding the trade-off between benefits (e.g. reduction in death or ESKD) and harms (e.g. increase in serious infections). The survey revealed : 1) Given a 6% increased risk of serious infections, most patients would want a benefit of at least 4.5% reduction in end-stage kidney disease in a time frame of 1 year (the median of the panel's votes) for plasma exchange; 2) if the risk increase in serious infections was 18% (instead of 6%), most patients would want a benefit of at least 7% decrease in risk of end-stage kidney disease (the median of the panel's votes) for plasma exchange. The panel concluded that patients put a higher value on end-stage kidney disease and a relative less value on serious infections.

Appendix 3 Second panel survey of values and preferences towards plasma exchange in patients with ANCA-associated vasculitis (AAV)

The absolute effects of plasma exchange on both reduction in end-stage kidney disease and increase in serious infections increase as patients' serum creatinine rises. The second survey asked panel members, based on their own experience, to make judgements about how patients within particular ranges of serum creatinine level would view the trade-off between benefit (i.e. reduction in end-stage kidney disease) and harm (i.e. increase in serious infections) of plasma exchange. Of the 22 panel members, 19 completed the online survey. One patient and one caregiver completed the questionnaire together, so we received 18 responses in total. The survey questionnaire and panel members' votes are as follows (number of votes given after each response alternative):

Introduction

Purpose of this survey: We would like to know the panel's perspective regarding your view about the distribution of choices individuals would make after full shared decision-making regarding whether or not to use plasma exchange. We will use your responses to inform our discussion of the tipping point, with regard to baseline creatinine levels, at which the majority would switch from declining to accepting plasma exchange.

Content of this survey: We will present to you 1) a summary of your views of patients' values and preferences towards plasma exchange from the first panel survey; 2) the benefits (absolute risk reduction in end-stage kidney disease) and harms (absolute risk increase in serious infections) of plasma exchange in patients with ANCA-associated vasculitis and with baseline creatinine at <200μmol/Lμmol/L, 200-300μmol/Lμmol/L, 300-400μmol/Lμmol/L, 400-500μmol/L, >500μmol/L. We will then ask you for your perspective about what proportion of patients would choose plasma exchange under each condition.

Each question will vary the risk reduction of end-stage kidney disease and the risk increase of serious infection. Please read these carefully.

Summary of values and preferences towards plasma exchange from the first panel survey

If a patient who takes plasma exchange has a **6 in 100 increased risk of serious infection** in the first year, the panel thinks the patient would require at least a **4.5 in 100 decreased risk of end-stage kidney disease**.

If a patient who takes plasma exchange has an **18 in 100 increased risk of serious infection** in the first year, the panel thinks the patient would require at least **a 7 in 100 decreased risk of end-stage kidney disease**

Second panel survey

1. For patients with ANCA-associated vasculitis and with **serum creatinine < 200μmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **0.4 in 100 lower risk** of end-stage kidney disease at 1 year

Harms: **2.7 in 100 increased risk** of serious infections at 1 year

- All or almost all (over 90%) would **choose** plasma exchange: 0
- Most (75- 90%) would **choose** plasma exchange: 4
- Majority (51-74%) would **choose** plasma exchange: 0
- Majority (51-74%) would **decline** plasma exchange: 1
- Most (75-90%) would **decline** plasma exchange: 6
- All or almost all (over 90%) would **decline** plasma exchange: 7

2. For patients with ANCA-associated vasculitis and with **serum creatinine > 500µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **16.8 in 100 lower risk** of end-stage kidney disease at 1 year
Harms: **13.5 in 100 increased risk** of serious infections at 1 year

- All or almost all (over 90%) would **choose** plasma exchange: 9
- Most (75- 90%) would **choose** plasma exchange: 6
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 0
- Most (75-90%) would **decline** plasma exchange: 0
- All or almost all (over 90%) would **decline** plasma exchange:0

3. For patients with ANCA-associated vasculitis and with **serum creatinine 200-300µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **3.1 in 100 lower risk** of end-stage kidney disease at 1 year
Harms: **4.9 in 100 increased risk** of serious infections at 1 year

- All or almost all (over 90%) would **choose** plasma exchange: 2
- Most (75- 90%) would **choose** plasma exchange: 2
- Majority (51-74%) would **choose** plasma exchange: 3
- Majority (51-74%) would **decline** plasma exchange: 8
- Most (75-90%) would **decline** plasma exchange: 2
- All or almost all (over 90%) would **decline** plasma exchange: 1

4. For patients with ANCA-associated vasculitis and with **serum creatinine 400-500µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **10.4 in 100 lower risk** of end-stage kidney disease at 1 year
Harms: **9.7 in 100 increased risk** of serious infections at 1 year

- All or almost all (over 90%) would **choose** plasma exchange: 5
- Most (75- 90%) would **choose** plasma exchange: 8
- Majority (51-74%) would **choose** plasma exchange: 4
- Majority (51-74%) would **decline** plasma exchange: 1
- Most (75-90%) would **decline** plasma exchange: 0

- All or almost all (over 90%) would **decline** plasma exchange: 0

5. For patients with ANCA-associated vasculitis and with **serum creatinine <u>300-400µmol/L</u>**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **5.7 in 100 lower risk** of end-stage kidney disease at 1 year
Harms: **7.3 in 100 increased risk** of serious infections at 1 year

- All or almost all (over 90%) would **choose** plasma exchange: 3
- Most (75- 90%) would **choose** plasma exchange: 1
- Majority (51-74%) would **choose** plasma exchange: 12
- Majority (51-74%) would **decline** plasma exchange: 1
- Most (75-90%) would **decline** plasma exchange: 1
- All or almost all (over 90%) would **decline** plasma exchange: 0

Conclusions

The majority of panel members think that the majority of patients with serum creatinine < 300 µmol/L would decline plasma exchange, while the majority of patients with serum creatine ≥ 300 µmol/L would choose plasma exchange.

Appendix 4 Rapid Recommendation panel members and their declaration of interests

Clinical Chair	
Alfred Mahr Rheumatologist	Rheumatology Clinic, Department of Internal Medicine, Kantonsspital St. Gallen, St. Gallen, Switzerland
Methods Co-Chair coaches	
Gordon Guyatt General internist, guideline expert, methodologist	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada ; Department of Medicine, McMaster University, Hamilton, Ontario, Canada
Reed Siemieniuk General internist, General internist, guideline expert, methodologist	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
Methods Co-Chair	
Linan Zeng Methodologist, pharmacist	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada ; Pharmacy department/Evidence-based pharmacy center, West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China
Clinical Experts	
Lynn Fussner Intensivist, pulmonologist	Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA
Karin Kilian Rheumatologist	Department of Rheumatology, Oslo University Hospital, Oslo, Norway; Institute of Clincal Medicine, University of Oslo, Oslo, Norway.
Mark Little Nephrologist	Trinity Translational Medicine Institute, Trinity College Dublin, Ireland ; Irish Centre for Vascular Biology, Tallaght University Hospital, Dublin, Ireland
Thomas Mavrakanas Nephrologist	Department of Medicine, Geneva University Hospitals and Faculty of Medicine, Geneva, Switzerland ; Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada
Reem A. Mustafa Nephrologist	Department of Internal Medicine, Division of Nephrology and Hypertension, University of Kansas Medical Center, Kansas, United States

Maryam Piram Pediatrician	CHU Sainte Justine Research Center, Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada ; CEREMAIA, Centre d'épidémiologie et de santé des populations (CESP), University Paris-Saclay, Le Kremlin Bicêtre, France
Lisa Katrina Stamp Rheumatologist	University of Otago Christchurch, , Christchurch, New Zealand
Michael Walsh Nephrologist	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada ; The Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada ; Department of Medicine, McMaster University, Hamilton, Ontario, Canada; St. Joseph's Healthcare, Hamilton, Ontario, Canada
Patient/Caregiver Partners	
Michelle Booth	United States
Paul Brown	United States
Mark Farrar	United Kingdom
Lesha Farrar	United Kingdom
Tracy Firth	United Kingdom
Methods Experts	
Thomas Agoritsas General internist, guideline expert, methodologist	Division of General Internal Medicine & Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland ; Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
David Collister Patient liaison, nephrologist	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada ; The Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada ; Department of Medicine, McMaster University, Hamilton, Ontario, Canada; St. Joseph's Healthcare, Hamilton, Ontario, Canada
Lyubov Lytvyn Patient liaison, methodologist	Department of Health Research Methods, Evidence and Impact, McMaster University,

	Hamilton, Ontario, Canada
Per O. Vandvik General internist, guideline expert, methodologist	Department of Medicine, Lovisenberg Hospital Trust, Oslo, Norway
Yingqi Xiao Methodologist of systematic review	Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada ; West China School of Nursing / Department of Nursing, West China Hospital, Sichuan University, China.

Pre-screening

All panel members were pre-screened for conflicts of interest prior to the guideline process that resulted in the *BMJ Rapid Recommendations*. The pre-screening was performed by the RapidRecs steering committee, affiliated with the non-profit organisation *MAGIC* (www.magicproject.org) and with support and approval from at least two unconflicted *BMJ* editors. No financial conflicts of interest were allowed (specifically, no financial ties to pharmaceutical companies with any stake in gastric acid suppressants) and intellectual and professional conflicts of interest were managed appropriately (see appendix 2: Methods for *BMJ* Rapid Recommendations). Panel members could not have a conflict for the past three years and do not anticipate a conflict arising in the foreseeable future, which we defined as at least one year.

Disclosures

Financial disclosures: No panel members had any financial conflicts of interest to disclose related to this clinical question.

Professional disclosures: Almost all of the physician panel members routinely see patients to whom this guideline applies, but their practice, rank, and remuneration will be unaffected by these recommendations.

Intellectual disclosures: Michael Walsh, David Collister, Gordon Guyatt, Alfred Mahr, Linan Zeng, Yingqi Xiao, and Reed Siemieniuk participated in writing the systematic review that formed the evidence base for this guideline. Michael Walsh participated in writing the prognostic study that support the stratification of recommendations in this guideline. Michael Walsh is the co-PI of PEXIVAS trial. Reed Siemieniuk, Thomas Agoritsas, Per Vandvik, Lyubov Lytvyn, Linan Zeng and Gordon Guyatt are members of the GRADE Working Group: *BMJ Rapid Recommendations* adheres to GRADE methods.

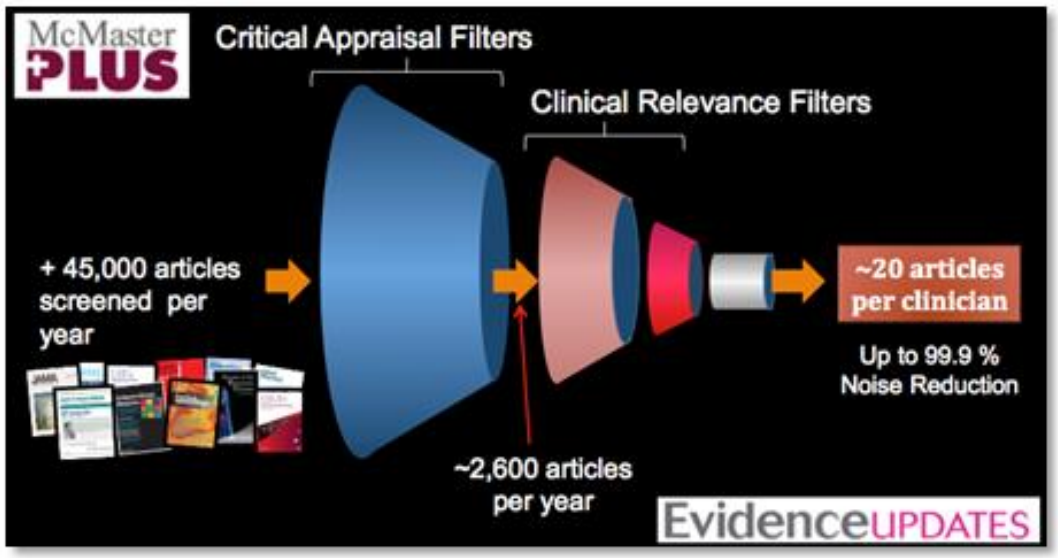
Appendix 5 Methodology for development of *BMJ Rapid Recommendations*

About *BMJ Rapid Recommendations*

Translating research to clinical practice is challenging. Trustworthy clinical practice recommendations are one useful knowledge translation strategy. Organisations creating systematic reviews and guidelines often struggle to deliver timely and trustworthy recommendations in response to potentially practice-changing evidence. *BMJ Rapid Recommendations* aims to create trustworthy clinical practice recommendations based on the highest quality evidence in record time. The project is supported by an international network of systematic review and guideline methodologists, people with lived experience of the diseases or conditions, clinical specialists, and front-line clinicians. This overview is one of a package that includes recommendations and one or more systematic reviews published by the *BMJ* group and in MAGICapp (http://www.magicapp.org). The goal is to translate evidence into recommendations for clinical practice in a timely and transparent way, minimizing bias and centred around the experience of patients. *BMJ Rapid Recommendations* will consider both new and old evidence that might alter established clinical practice.

Process overview

1. On a daily basis, we monitor the literature for practice-changing evidence:
   a. Formal monitoring through McMaster Premium LiteratUre Service (PLUS)



   b. Informal monitoring the literature by *BMJ Rapid Recommendations* expert groups, including clinician specialists and patients
2. The *RapidRecs* executive team and editors at *The BMJ* choose which clinical questions to pursue among the identified potentially-practice changing evidence, based on relevance to a wide audience, widespread interest, and likelihood to change practice.

3. We incorporate the evidence into the existing body of evidence and broader context of clinical practice via:

    a. a rapid and high-quality systematic review and meta-analysis on the benefits and harms with a focus on the outcomes that matter to patients

    b. parallel rapid recommendations that meet the standards for trustworthy guidelines[1] by an international panel of people with relevant lived experience, front-line clinicians, clinical content experts, and methodologists.

    c. The systematic review and the recommendation panel will apply standards for trustworthy guidelines.[1,2] They use the GRADE approach, which has developed a transparent process to rate the quality (or certainty) of evidence and grade the strength of recommendations.[3,4]

    d. Further research may be conducted including:

        i. A systematic review of observational studies to identify baseline risk estimates that most closely represent the population at the heart of the clinical question, a key component when calculating the estimates of absolute effects of the intervention

        ii. A systematic review on the preferences and values of patients on the topic.

4. Disseminate the rapid recommendations through

    a. publication of the research in *BMJ* journals

    b. short summary of recommendations for clinicians published in *The BMJ*

    c. press release and/or marketing to media outlets and relevant parties such as patient groups

    d. Links to BMJ Group's *Best Practice* point of care resource

    e. MAGICapp which provides recommendations and all underlying content in digitally structured multilayered formats for clinicians and others who wish to re-examine or consider national or local adaptation of the recommendations.

Who is involved?

Researchers, systematic review and guideline authors, clinicians, and patients often work in silos. Academic journals may publish work from any one or combinations of these groups of people and findings may also be published in the media. But it is rare that these groups work together to produce a comprehensive package. *BMJ-RapidRecs* circumvents organisational barriers in order to provide clinicians with guidance for potentially practice-changing evidence.

Our collaboration involves

    a. The *RapidRecs* group with a designated Executive team responsible for recruiting and coordinating the network of researchers who perform the systematic reviews and the recommendation panels.. The *RapidRecs* group is part of *MAGIC* (www.magicproject.org), a non for profit organization that provides MAGICapp (www.magicapp.org) an authoring and publication platform for evidence

summaries, guidelines and decision aids, which are disseminated online for all devices.[5]

b. *The BMJ* helps identifiying practice-changing evidence on key clinical questions, coordinates the editorial process and publishes the package of content linking to the MAGICapp that is presented in a user friendly way.

METHODS FOR THE RAPID RECOMMENDATIONS

The formation of these recommendations adheres to standards for trustworthy guidelines with an emphasis on patient involvement, strict management of conflicts of interests, as well as transparent and systematic processes for assessing the quality of evidence and for moving from evidence to recommendations.[1,2,6]

Guidance on how the panel is picked and how they contribute

Panel members are sought and screened through an informal process.

The following panel members are important

- At least one author of the individual systematic reviews
- At least one patient representative with lived experience of the disease or condition. This person receives patient-oriented documents to explain the process and is allocated a linked panel member to empower their contribution.
- A full spectrum of practicing clinicians involved in the management of the clinical problem and patients it affects, including front-line clinicians with generalist experience and those with deep content clinical and research expertise in the particular topic.
- Methodological experts in health research methodology and guideline development

Any potential conflicts of interest are managed with extreme prudence:

- No panel member can have a financial interest – as assessed by the panel chair, the *Rapidrecs* executive team or *The BMJ* editors as relevant to the topic
- No more than two panel members with an intellectual interest on the topic (typically having published statements favouring one of the interventions).

<u>Illustrative example:</u> For the BMJ Rapid Recommendations on antiretroviral therapy for pregnant women living with HIV, the panel recruitment of content experts and community panel members was challenging. Content experts in this area are infectious diseases experts, many of whom have financial conflicts of interests through interactions with the pharmaceutical industry through advisory boards and participation in industry-funded trials. The group reached out to more than 17 potential panel members who were eventually excluded from participating because of conflicts – notably, all of these persons had not disclosed any relevant conflicts on related and recent publications in the topic area. Many more potential panel members were not recruited because of publicly declared conflicts. The chair and MAGIC team were able, with considerable effort and ingenuity, to recruit several excellent and unconflicted content experts.

How the panel meets and works

The international panel communicates via teleconferences and e-mail exchange of written documents throughout the process. Minutes from teleconferences are audiorecorded, transcribed, and stored for later documentation (available for peer-reviewers on request).

Teleconferences typically occur at three timepoints, with circulated documents by e-mail in advance:

1. At the initiation of the process to provide feedback on the systematic review protocol (for example, on selection of patient-important outcomes and appropriate prespecified analysis of results) before it is performed.
2. At the evidence summary stage with discussion, feedback and agreement on draft evidence (GRADE evidence profile) prepared by the Chair and the methods editor based on the systematic review.
3. At the recommendation formulation phase with discussion, feedback and agreement on draft recommendations and other content underlying the recommendation (e.g. GRADE SoF-table, key information, rationale, practical advice)

Following the last teleconference the final version of the recommendations is circulated by e-mail specifically requesting feedback from all panel members to document agreement before submission to *The BMJ*. Additional teleconferences are arranged as needed.

Illustrative example: For the BMJ Rapid Recommendations on antiretroviral therapy for pregnant women living with HIV, two large-group teleconferences were arranged. First, content experts provided crucial input to evidence assessment (e.g. subgroups to identify). For the recommendation formulation phase the panel needed two teleconferences to discuss all elements in detail, followed by more than 100 e-mails with specific issues to be sorted out. Multiple teleconferences were held to allow the scheduling flexibility required so that all could participate.

How we move from research findings to recommendations

What information is considered?

The panel considers best current evidence from available research. Beyond systematic reviews - performed in the context of the *BMJ Rapid Recommendations* - the panel may also include a number of other research papers to further inform the recommendations.

How is a trustworthy guideline made?

The Institute of Medicine (IOM)'s guidance on out how trustworthy guidelines should be developed and articulated key standards as outlined in the table below.[1] The standards are similar to those developed by the Guideline International Network (G-I-N).[2] These standards have been widely adopted by the international guideline community. Peer reviewers of the recommendation article are asked whether they found the guideline trustworthy (in accordance

with IOM standards). The table below lays out how we hope to meet the standards for our rapid recommendations:

1. Establishing transparency **"The processes by which a CPG is developed and funded should be detailed explicitly and publicly accessible"***
● This method is available and published as a supplementary file as well as in MAGICapp where all recommendations and underlying content is available. ● We ask the peer-reviewers to judge whether the guidance is trustworthy and will respond to concerns raised.
2. Managing conflicts of interest **"Prior to selection of the guideline development group, individuals being considered for membership should declare all interests and activities potentially resulting in COI with development group activity....",**
● Interests of each panel member are declared prior to involvement and published with the rapid recommendations ● No one with any potential financial interests in the past three years, or forthcoming 12 months will participate - as judged by the panel chair and *The BMJ* ● No more than two panel members have declared an intellectual conflict of interest. Such conflicts include having taken a position on the issue for example by a written an editorial, commentary, or conflicts related to performing a primary research study or written a prior systematic review on the topic. ● The Chair must have methods expertise, a clinical background and no financial or intellectual interests. ● Funders and pharmaceutical companies have no role in these recommendations.
3. Guideline Development Group Composition **"The guideline development group should be multidisciplinary and balanced, comprising a variety of methodological experts and clinicians, and populations expected to be affected by the CPG"**
● *The RapidRecs* group will aim to include representation from most or every major geographic region in the world, with specific efforts made to achieve gender-balance.

- We will facilitate patient and public involvement by including patient experience, via patient-representatives and systematic reviews addressing values and preferences to guide outcome choices and relative weights of each outcome, where available
- Patient-representatives will be given priority during panel meetings and will have an explicit role in vetting the panel's judgements of values and preferences.

4. Clinical Practice Guideline–Systematic Review Intersection

"CPG developers should use systematic reviews that meet standards set by the IOM. Guideline development group and systematic review team should interact regarding the scope, approach, and output of both processes".

- Each rapid recommendation will be based on one or more high-quality SRs either developed and published in parallel with our *BMJ Rapid Recommendations* or produced by other authors and available at the time of making the recommendaiton.
- The recommendation panel and SR teams will interact, with up to three members participating in both teams to facilitate communication and continuity in the process

5. Establishing Evidence Foundations for and Rating Strength of Recommendations

"For each recommendation: explain underlying reasoning, including a clear description of potential benefits and harms, a summary of relevant available evidence and description of the quality., explain the part played by values, opinion, theory, and clinical experience in deriving the recommendation, "provide rating of strength of recommendations"

- The GRADE approach will provide the framework for establishing evidence foundations and rating strength of recommendations.[6] For each recommendation systematic and transparent assessments are made across the following key factors:
  - Absolute benefit and harms for all patient-important outcomes through structured evidence summaries (e.g. GRADE Summary of Findings tables)[4]
  - Quality of the evidence[7]
  - Values and preferences of patients
  - Resources and other considerations (e.g. feasibility, applicability, equity)
- Each outcome will - if data are available through systematic reviews - include an effect estimate and confidence interval, with a measure of certainty in the evidence, as presented in Summary of Findings tables. If such data are not available narrative summaries will be provided.
- A summary of the underlying reasoning and all additional information (e.g. key factors, practical advice, references) will be available online in an interactive format at www.magicapp.org. This summary will include descriptions of how theory (e.g.

patophysiology) and clinical experience played into the evidence assessment and recommendation development.

- Recommendations will be rated either weak or strong, as defined by GRADE.[8]
- If the panel members disagree regarding evidence assessment or strength of recommendations, we will follow a structured consensus process customized to the GRADE system and report any final differences in opinion, with their rationale, in the online supplement and online at www.magicapp.org.

6. Articulation of recommendations

"Recommendations should be articulated in a standardized form detailing precisely what the recommended action is, and under what circumstances it should be performed, and so that compliance with the recommendation(s) can be evaluated"

- Each recommendation will appear at the top of the guideline infographic, published in *The BMJ*, and will be available in standardised formats in MAGICapp, articulated to be actionable based on best current evidence on presentation formats of guidelines.[9]
- There will be a statement included in each summary article in *The BMJ* and in the MAGICapp that these are recommendations to provide clinicians with guidance. They do not form a mandate of action and should be contextualised in the healthcare system a clinician's works in, and or with an individual patient.

7. External review

"External reviewers should comprise a full spectrum of relevant stakeholders...., authorship should be kept confidential....., all reviewer comments should be considered....a rationale for modifying or not should be recorded in writing.... a draft of the recommendation should be made available to general public for comment.."

- At least two external peer-reviewers and one patient reviewer will review the article for *The BMJ* and provide open peer review. Each will have access to all the information in the package. They will be asked for general feedback as well as to make an overall judgement on whether they view the guidelines as trustworthy
- A *BMJ* series adviser with methodological and/or statistical expertise will review the *BMJ* Rapid Recommendations publication and the systematic reviews.
- The *Rapidrecs* panel will be asked to read and respond to the peer review comments and make amendments where they judge reasonable
- *The BMJ* and *RapidRecs* executive team may, on a case-by-case basis, choose to

invite key organizations, agencies, or patient/public representatives to provide and submit public peer-review.

- There will be post-publication public review process through which people can provide comments and feedback through MAGICapp (or through *The BMJ*). The Chair will, on behalf of panel authors, aim to respond to each publicly-available peer-review within 30 days, for a period of six months after publication.

8. Updating

"The date for publication, systematic review and proposed date for future review should be documented, the literature should be monitored regularly and the recommendation should be updated when warranted by new evidence"

- The *Rapidrecs* panel will, through monitoring of new research evidence for published *BMJ Rapid Recommendations,* aim to provide updates of the recommendations in situations in which the evidence suggests a change in practice. These updates will be initially performed in MAGICapp and submitted to *The BMJ* for consideration of publication of a new Rapid Recommendation.

References:

1. Laine C, Taichman DB, Mulrow C. Trustworthy clinical guidelines. *Ann Intern Med. 2011;154(11):774-775. doi:10.7326/0003-4819-154-11-201106070-00011*
2. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med*. 2012;156(7):525-531. doi:10.7326/0003-4819-156-7-201204030-00009
3. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *BMJ. 2008;336(7652):1049-1051.* doi:10.1136/bmj.39493.646875.AE
4. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol. 2011;64(4):383-394.* doi:10.1016/j.jclinepi.2010.04.026
5. Vandvik PO, Brandt L, Alonso-Coello P, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. Chest. 2013;144(2):381-389. doi:10.1378/chest.13-0746
6. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD
7. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-406. doi:10.1016/j.jclinepi.2010.07.015
8. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol*. 2013;66(7):726-735. doi:10.1016/j.jclinepi.2013.02.003.

9. Kristiansen A, Brandt L, Alonso-Coello P, et al. Development of a novel, multilayered presentation format for clinical practice guidelines. *Chest*. 2015;147(3):754-763. doi:10.1378/chest.14-1366

Chapter 3

GRADE guideline 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings

Zeng L, Brignardello-Petersen R, Hultcrantz M, Siemieniuk RAC, Santesso N, Traversy G, Izcovich A, Sadeghirad B, Alexander PE, Devji T, Rochwerg B, Murad MH, Morgan R, Christensen R, Schünemann HJ, Guyatt GH

Journal of Clinical Epidemiology; 2021; 137:163-175.

Abstract

Objective —— To provide practical principles and examples to help systematic review and health technology assessment authors make optimal choices regarding their ratings of certainty of evidence using a minimally or partially contextualized approach.

Study Design and Setting —— Based on the GRADE clarification of certainty of evidence in 2017, a project group within the GRADE Working Group conducted iterative discussions and presentations at GRADE Working Group meetings to refine this construct and produce practical guidance.

Results —— Systematic review and health technology assessment authors need to clarify what it is in which they are rating their certainty of evidence (i.e. the target of their certainty rating). The decision depends on the degree of contextualization (partially or minimally contextualized), thresholds (null, small, moderate or large effect threshold), and where the point estimate lies in relation to the chosen threshold(s). When the 95% confidence interval crosses multiple possible thresholds (e.g. including both large benefit and large harm), it is not worthwhile for authors to determine the target of certainty rating.

Conclusion —— GRADE provides practical principles to help systematic review and health technology assessment authors specify the target of their certainty of evidence rating.

Keywords

GRADE; target of certainty of evidence rating; thresholds; Evidence-based medicine; systematic review; health technology assessment

Running title

Guidance on choosing the target of GRADE certainty of evidence ratings

What is new?

Key findings

- Systematic review and health technology assessment authors need to determine the target when they make GRADE ratings of certainty of evidence.
- This decision depends on the degree of contextualization, the threshold(s) chosen, and the relative position of point estimate in relation to the chosen threshold(s).
- When 95% confidence interval crosses multiple possible thresholds (e.g. including both large benefit and large harm), it is not worthwhile to determine the target of the certainty rating.

What this adds to what was known

- Building on prior GRADE guidance, this article provides specific suggestions for deciding on the target of certainty of evidence ratings.
- We provide practical guidance on how to make optimal choices regarding rating certainty of evidence using a minimally or partially contextualized approach.

What is the implication and what should change now

- The article will help systematic review and health technology assessment authors be aware of the importance of determining the target of their rating of certainty of evidence when using GRADE.

- Whenever they rate the certainty of evidence, systematic review and health technology assessment authors should be explicit about the target of the rating.

1. Background

GRADE Guidance thus far

In previous guidance for authors of systematic reviews, health technology assessments, and clinical practice guidelines, the GRADE working group has offered clarification regarding how to make ratings of certainty (quality) of bodies of evidence [1]. In particular, when considering a choice between candidate interventions, ratings of certainty of evidence reflect our confidence that the true effect of an outcome lies on one side of a threshold (e.g. on the left side of the small effect threshold in Figure 1) or within a chosen range (e.g. within the range of small effect in Figure 1). In this article, written for systematic review and health technology assessment authors, we further clarify previous guidance.

In dealing with how to address the threshold or range, GRADE notes the importance of deciding on the associated level of contextualization. The choice of level of contextualization depends on what audiences would find most useful. In particular, for systematic reviews and health technology assessments, if audiences' focus is on whether there is an effect or whether there is an important effect, authors would choose a minimally contextualized approach. If audiences' focus is on the magnitude of effect (i.e. a trivial, small, moderate or large), authors would use a partially contextualized approach. Appendix 1 presents further semantic and conceptual issues related to minimally contextualized and partially contextualized approaches.

The minimally contextualized approach specifies two possible thresholds: no difference between groups (e.g. risk ratio [RR] of 1.0, risk difference [RD] of 0) or an important effect (i.e. minimal important difference [MID], also called small effect threshold) as a threshold for rating the certainty. Using a partially contextualized approach, reviewers would rate their certainty in relation to a range (i.e. a range of trivial, small, moderate or large effect) that is bounded by two thresholds (Figure 1) [1-4].

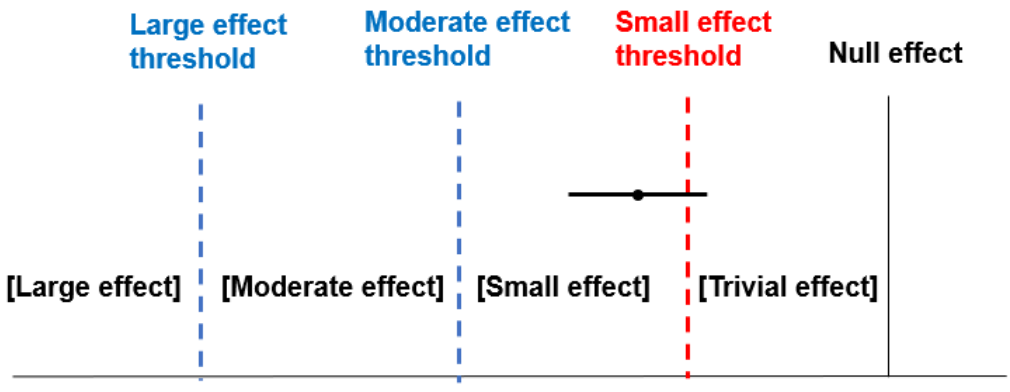


Figure 1 Rating of certainty of evidence using a minimally or partially contextualized approach
Null effect and small effect threshold are possible thresholds for a minimally contextualized approach, while small, moderate and large effect thresholds provide the boundaries of ranges (i.e. a range of trivial, small, moderate or large effect) for a partially contextualized approach. A small effect threshold is also called a minimally important difference (MID).

Where is practical guidance still needed?

GRADE users still often fail to make an explicit statement about what it is in which they are rating their certainty (i.e., the target of the rating of certainty of evidence). This failure can have important consequences, because the judgments that influence the rating might depend on the choice of target.

For instance, consider a situation in which the risk difference (RD) in mortality between intervention A and placebo is 2 fewer deaths per 100 patients, with a 95% confidence interval (CI) from 0.5 fewer to 4 fewer death per 100 patients (Figure 2). Some may rate their certainty that intervention A reduces mortality when compared to placebo (i.e. the target of certainty rating) and thus require no rating down for imprecision. Others might rate their certainty that there is an important difference in mortality between intervention A and placebo (i.e. the target of certainty rating). If they set the small effect threshold at 1% reduction of mortality, they would rate down for imprecision. Others may believe that users of their systematic review would be optimally informed by seeing both ratings.

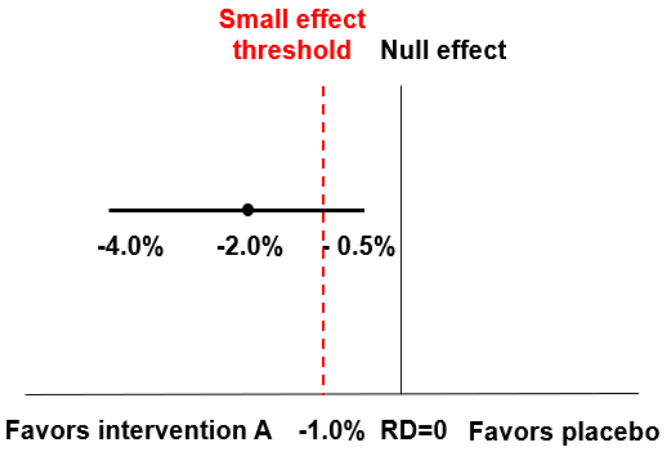


Figure 2 A hypothetic example: intervention A versus placebo in reduction of mortality

Some may rate their certainty that intervention A reduces mortality when compared to placebo and not rate down for imprecision; Others may rate their certainty that intervention A has an important reduction in mortality compared with placebo and – depending on their threshold from importance, 1%, for example – would rate down for imprecision.

Scope of this article

This article provides practical principles and examples to help systematic review and health technology assessment authors make optimal choices regarding the target of their rating of certainty of evidence using a minimally or partially contextualized approach. In systematic reviews and health technology assessments, authors want to learn about the effect of interventions, and consider one outcome at one time. This article does not address issues that arise using a fully contextualized approach (e.g., a clinical practice guideline) or how the ratings would affect recommendations [5].

Further, being explicit about the basis for a threshold is important. However,
the article will not address how to set thresholds other than the null. Choice of threshold will depend on the perspective (e.g., clinical perspective, public health perspective), the context (e.g. settings with more or less well-developed health care resources), and patients' values and preferences (including the importance of the outcomes), among other factors.

2. Practical principles for deciding about the target of rating of certainty of evidence

We suggest four principles for choosing the target of rating the certainty of evidence and describe and illustrate these principles below. For simplification, we focus on single paired comparisons, and restrict the examples to situations in which results suggest a reduction in harmful outcomes (i.e. a reduction in the occurrence of the outcome is desirable). These principles, however, can be applied to ratings of certainty of evidence in all situations.

Principle 1 *Reviewers need to decide about the target of their certainty ratings*.
Given that one could rate one's certainty that the true effect lies on one or the other side of a threshold, or that it lies within a particular range (i.e. between two thresholds) (Figure 1) [1], GRADE users should be explicit regarding the target of their certainty ratings.

In the hypothetic example of intervention A versus placebo on reduction of mortality (Figure 2), when rating certainty in relation to the null effect, reviewers could specify they are rating their certainty that the true effect is a non-null effect (i.e., there is an effect), in this case a reduction in mortality greater than zero. Alternatively, when rating certainty in relation to the small effect threshold, they could specify they are rating their certainty that the true effect is important, in this case a reduction in mortality greater than the small effect threshold.

Principle 2 *The target of certainty ratings will depend on the degree of contextualization, the threshold chosen, and the point estimate.*

a. Degree of contextualization

The decision regarding the target of certainty rating will differ depending on the degree of contextualization. For systematic reviews and health technology assessment authors, both minimally and partially contextualized approaches prove practical and useful.

b. The threshold choice

When using a minimally contextualized approach, reviewers most often rate their certainty in relation to a single threshold. The null effect or a small effect threshold represent possible thresholds (Figure 1) [1]. Rating certainty in relation to the former results in a rating of certainty in whether a non-null effect is present, while rating certainty in relation to the latter leads to a rating of certainty in whether an important effect is present.

When using a partially contextualized approach, one makes ratings in relation to a range. GRADE suggests four possible ranges (i.e., a range of trivial, small, moderate, or large effect) divided by three thresholds (i.e., small, moderate or large effect threshold) (Figure 1) [1]. Choosing to rate certainty in relation to a particular range would result in a rating of certainty in whether a particular magnitude of effect is present.

When rating the certainty in relation to the null effect, one can present the threshold (i.e. the null effect) and effect estimates in either relative (e.g., RR =1) or absolute terms (e.g., RD=0). Depending on baseline risks, however, a particular relative effect may correspond with very different absolute effects. Thus, rating certainty in relation to threshold(s) other than the null effect requires presenting both threshold(s) and effect estimates in absolute terms [6]. Box 1 clarifies calculation of absolute risks from relative risks and baseline risks.

Systematic review and health technology assessment authors can enhance transparency by reporting, in pre-registered protocols, the degree of contextualization and the particular threshold(s) or range(s) they will consider. For the threshold setting, as any threshold will involve some degree of uncertainty, authors could specify a range within which the threshold is likely to lie. Doing so, however, is likely to add considerable complexity to judgements, and is not something we would currently suggest.

Box 1 GRADE's approach of calculating an absolute risk from relative risk and baseline risk

c. The point estimate

We could calculate an absolute risk difference (RD) of an intervention versus a comparator using a risk ratio (RR) and an estimate of baseline risk (BR), with the following formula:

$$RD = BR\ (RR-1)$$

The baseline risk (BR) is the event rate in the comparator group (ranging from 0.00 to 1.00). We could obtain an estimated BR from population based observational studies or control groups of randomized controlled trials (RCTs) [7].

A RR below 1 represents a reduction in the risk of the event due to the intervention, and the RD could be presented as a negative number. A RR greater than 1 represents an increase in the risk of the event due to the intervention, and the RD would then be presented as positive. Alternatively, if one decided to frame the effect as an absolute risk reduction, the RD could be presented as positive.

We can also estimate the RD from an odds ratio (OR) with an estimate of BR, using the following formula [8]:

$$RD = [(1-OR)\ BR\ /\ \{1 - BR + (1-OR)\ BR\}]$$

For a given threshold, where the point estimate falls in relation to the chosen threshold(s) also determines the target of certainty rating. Figure 3 depicts the implications of particular point estimates when one chooses to rate certainty in relation to a small effect threshold. In situation (a), because the point estimate falls above the small effect threshold, one would rate one's certainty that the true effect is an important effect. In situation (b), however, as the point estimate falls below that threshold, one would rate one's certainty that the true effect is trivial or not

important (i.e. smaller than the small effect threshold).

Still using this example (Figure 3), one could choose a partially contextualized approach and rate certainty regarding whether a small effect is present (i.e. whether the true effect lies between small effect threshold and moderate effect threshold). In situation (a), because the point estimate falls within that range, one would rate certainty that the true effect is a small effect. In situation (b), as the point estimate falls below that range, one would rate certainty that the true effect is a trivial effect.

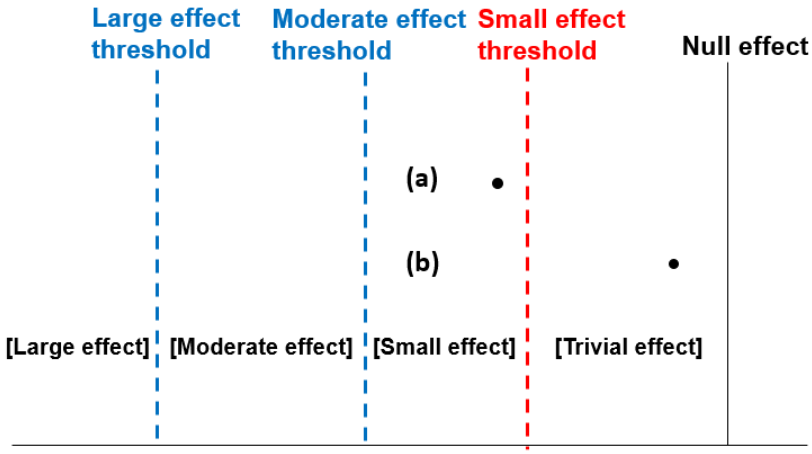


Figure 3 The location of point estimate in relation to the chosen threshold(s) would determine the target of certainty rating (Principle 2c)

(a) Reviewers would rate certainty that the true effect falls above the small effect threshold (i.e. an important effect is present) or within the range of small effect (i.e. a small effect is present); (b) Reviewers would rate certainty that the true effect falls below the small effect threshold or below the range of small effect (i.e. a trivial effect is present).

When the point estimate is very close to the chosen threshold, one approach (Approach 1) would rate the certainty that the true effect is either above or below that threshold, depending on which side the point estimate falls. However close the point estimate is to the threshold, it will always be possible to carry the calculation to as many decimal places as necessary to determine on which side of the threshold the point estimate lies. Another approach (Approach 2) would rate certainty in relation to adjacent threshold(s). We clarify these two approaches using the following hypothetical example.

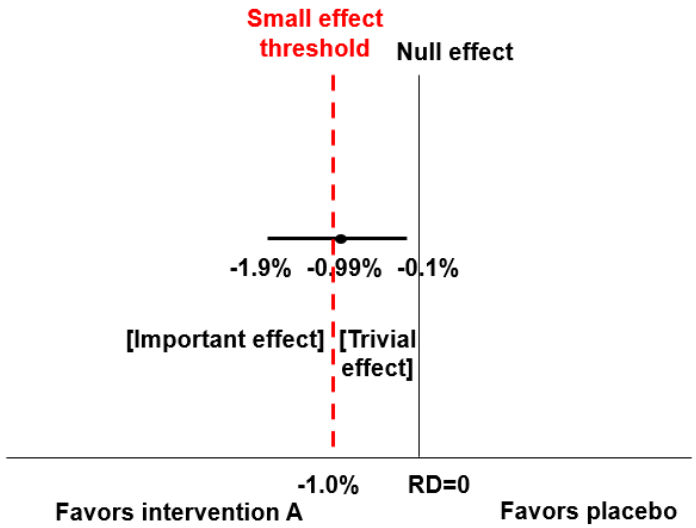
Consider a systematic review of intervention A versus placebo for prevention of stroke. Reviewers could had set a small effect threshold at 1 fewer stroke per 100 patients (Figure 4a). The meta-analysis yields a RD of 0.99 fewer strokes per 100 patients, with a 95% CI from 0.1 fewer to 1.9 fewer strokes per 100 patients.

Approach 1: Although the point estimate is very close to the chosen threshold (i.e., the small effect

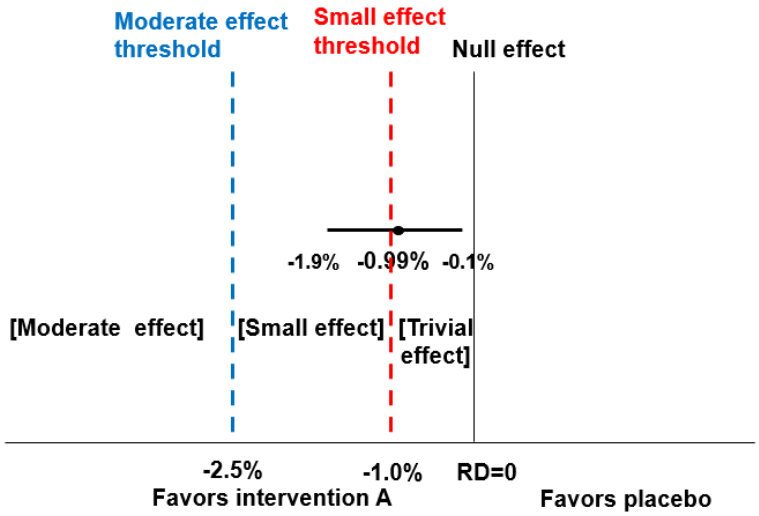
threshold), reviewers could still rate certainty that the true effect lies below that threshold. In this case, the reviewer would rate their certainty that the effect is trivial (Figure 4a).

Approach 2: Alternatively, reviewers could rate certainty in relation to two adjacent thresholds (i.e. the null effect, and a moderate effect threshold) (Figure 4b). As the point estimate falls within the two thresholds, they would rate certainty that the true effect is a trivial or small effect.

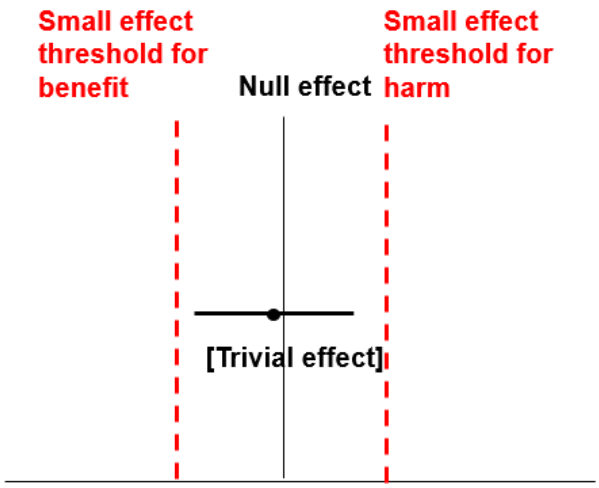
Reviewers might be more comfortable with applying Approach 2 to situations in which the point estimate lies at, or is very close to, the null effect. It is not possible to rate certainty in point estimates alone (i.e. that there is no effect). Instead, one would rate certainty in relation to a range of trivial effect between a small effect threshold for benefit and a small effect threshold for harm (Figure 4c). Because the point estimate falls within the range of trivial effect, reviewers would rate certainty that the true effect is a trivial to null effect.



(a) Still rating certainty rating in relation to the chosen threshold (i.e. the true effect is trivial).



(b) Rate certainty in relation to two adjacent thresholds (i.e. the true effect is trivial or small).



(c) Rate certainty in relation to two adjacent thresholds (i.e. the true effect is trivial to null).

Figure 4 Options for determining the target of the certainty rating when the point estimate is very close to the chosen threshold

Principle 3 *Using a particular degree of contextualization, where the reviewers set the threshold(s) will determine the target of the certainty rating.*

As presented in Figure 5, using a minimally contextualized approach and rating certainty in relation to a small effect threshold, if reviewers choose threshold 1, they would rate certainty that the true effect is larger than the small effect threshold (i.e. the true effect is an important effect). If they set the threshold at threshold 2, they would rate certainty that the true effect is smaller than the small effect threshold (i.e. the true effect is trivial).

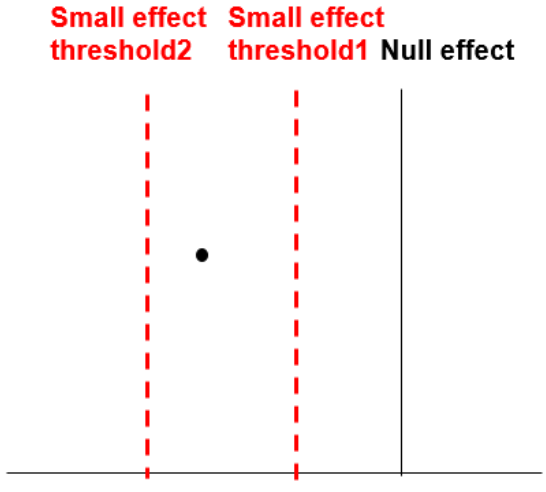


Figure 5 Where reviewers set the threshold(s) will determine the target of the certainty rating (Principle 3)

When choosing threshold 1, reviewers would rate certainty that the true effect falls above that threshold; when choosing threshold 2, they would rate certainty that the true effect falls below

that threshold.

Principles 1 to 3 represent fundamental rules for deciding the target of certainty of evidence rating. There is, however, an exception. Consider Figure 6 (a). Here, considering only the point estimate, one could rate certainty that the true effect is greater than the small effect threshold. Because the 95% CI is so wide that we have very little idea of where the true effect lies, this would, however, make little sense. The true effect might represent a very large benefit, or a very large harm, thus, Principle 4:

Principle 4 *When the 95% CI crosses multiple possible thresholds, it is not worthwhile to choose a particular threshold and hence not worthwhile to decide about the target of the rating of certainty of evidence.*

Under these circumstances (Figure 6(a)), rather than rating the certainty of evidence in relation to particular threshold(s), an appropriate conclusion would be reviewers have little idea of the true effect. Reviewers might also make this conclusion if the 95% CI included a large benefit and a small harm (Figure 6 (b)). In such situations, reviewers would rate down certainty of evidence by at least two levels. Exactly how wide the 95% CI has to be before reviewers abandon being explicit about the target of the certainty rating remains a matter of judgment.

In a partially contextualized approach one may (or may not) have specified boundaries of large benefit and large harm at the outset; in the minimally contextualized approach one would not. This does not preclude applying this principle in the minimally contextualized approach. Reviewers can still make intuitive judgments regarding how wide is too wide in terms of large benefit and harm.

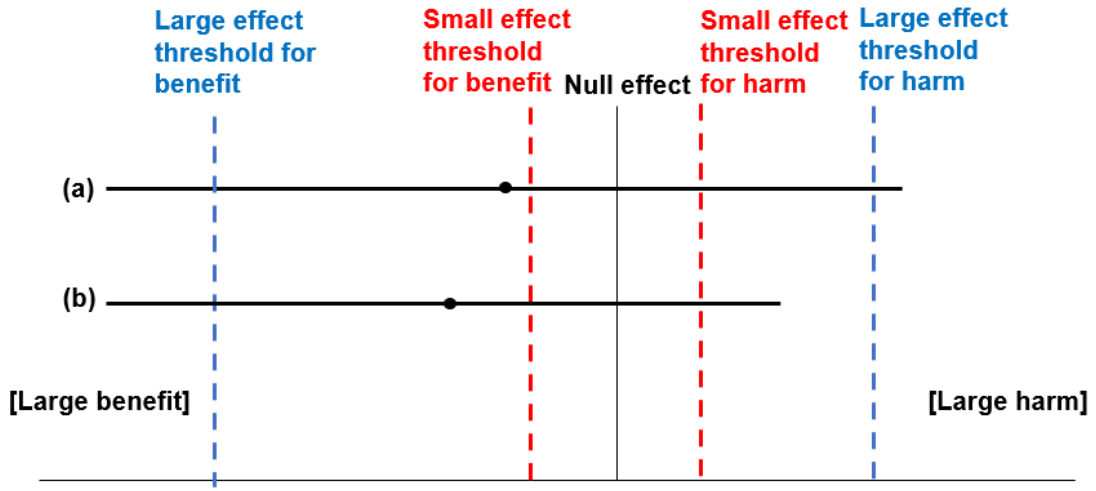


Figure 6 When 95% CI is extremely wide, it is not worthwhile to decide about the target of the rating of certainty of evidence. Reviewers are very uncertain where the true effect lies (Principle 4)

(a) The 95% CI crosses both threshold for large benefit and threshold for large harm; (b) The 95% CI crosses the threshold for large benefit and the threshold for small harm.

3. Application of principles

In this section, we describe how the principles presented above influence the judgments when rating the certainty of evidence. We focus on a single GRADE domain —— imprecision, assuming no serious limitation in the other four domains (i.e. risk of bias, indirectness, inconsistency, and publication bias). Appendix 2 presents further discussion regarding how one could consider the target of the rating of certainty of evidence, when simultaneously considering limitations in other GRADE domains [4,9].

Consider the hypothetical evidence in Figure 1.

1) If reviewers are interested in the certainty of whether there is an effect, they would choose a minimally contextualized approach and rate certainty in relation to the null effect. As the point estimate falls above that threshold, they would rate certainty that there is an effect. Because the 95% CI does not cross the null effect, they would not rate down for imprecision.

   Using the alternative within the minimally contextualized approach, reviewers would determine their certainty regarding whether an important effect exists, and would make their rating in relation to the small effect threshold. As the point estimate falls above that threshold, they would rate certainty that the effect is a small effect and would rate down for imprecision because the 95% CI crosses the chosen threshold.

2) If interest lies in the certainty of whether a small effect exists, using a partially contextualized approach, reviewers would rate certainty in relation to the small effect range. In this case, because the point estimate falls within that range, the rating would be in relation to a small effect. As the 95% CI crosses the small effect threshold and it therefore remains plausible that the effect is trivial, one would rate down for imprecision.

3) Using a partially contextualized approach, reviewers could rate the certainty that the effect is either small or trivial (i.e. it is smaller than moderate). If so, they would not rate down for imprecision because the 95% CI excludes values above the moderate effect threshold.

The chosen degree of contextualization and threshold(s) will depend on the audience. After review authors provide their rating, users can adjust ratings according to their own thresholds. For example, review authors could provide a rating that there is a benefit and not rate down for imprecision if the 95% CI excludes the null effect. A user might then make one's own rating in whether the true effect is small and rate down for imprecision if the 95% CI crosses the small effect threshold, or that the true effect is smaller than a large effect and not rate down for imprecision if the 95% CI does not cross the large effect threshold.

4. Real examples

We present two examples from a published systematic review [10] to illustrate the application of the principles. Appendix 3 presents more examples from published systematic reviews to illustrate the application of principles in this paper for continuous outcome, situation in which results suggest an increase in harmful outcomes and when the 95%CI is very wide. The authors of these reviews did not always follow the guidance we suggest here – thus illustrating the desirability of the

clarifications we present in this article.

Example 1: Determining the target of certainty rating using a minimally contextualized approach
Consider a systematic review of corticosteroids for patients with sepsis. A meta- analysis of 36 randomized controlled trials (RCTs) including 9,433 patients shows that corticosteroids yields 1.8 fewer deaths per 100 patients, with a 95% CI from 4.1 fewer to 0.8 more deaths per 100 patients (Figure 7) [10]. Using a minimally contextualized approach and rating certainty in relation to the null effect, the authors of the review could have rated their certainty that corticosteroids reduced mortality (i.e. there is an effect). Because the 95% CI crosses the null effect, they would then have rated down the certainty due to imprecision [10].

Alternatively, still using a minimally contextualized approach, the authors could have rated their certainty in relation to a small effect threshold. For example, had they set the small effect threshold at 0.5 fewer death per 100 patients, they could have rated their certainty that corticosteroids result in an important reduction in mortality. If, however, they had chosen mortality reduction of 2 per 100 patients as a small effect threshold, they could have rated their certainty that corticosteroids have a trivial reduction in mortality. Wherever they set the threshold, they would have rated down for imprecision due to the overlap of the 95% CI with the small effect threshold.

Example 2: Determining the target of certainty rating using a partially contextualized approach
We continue with the systematic review of corticosteroids for treatment of patients with sepsis, this time using a partially contextualized approach. The authors might have set a small effect threshold at 0.5 fewer deaths per 100 patients, and a moderate effect threshold at 3 fewer deaths per 100 patients (Figure 8). The authors could have then rated their certainty that corticosteroids result in a small reduction in mortality and rated down for imprecision because the 95% CI crosses both the small effect and moderate effect thresholds.

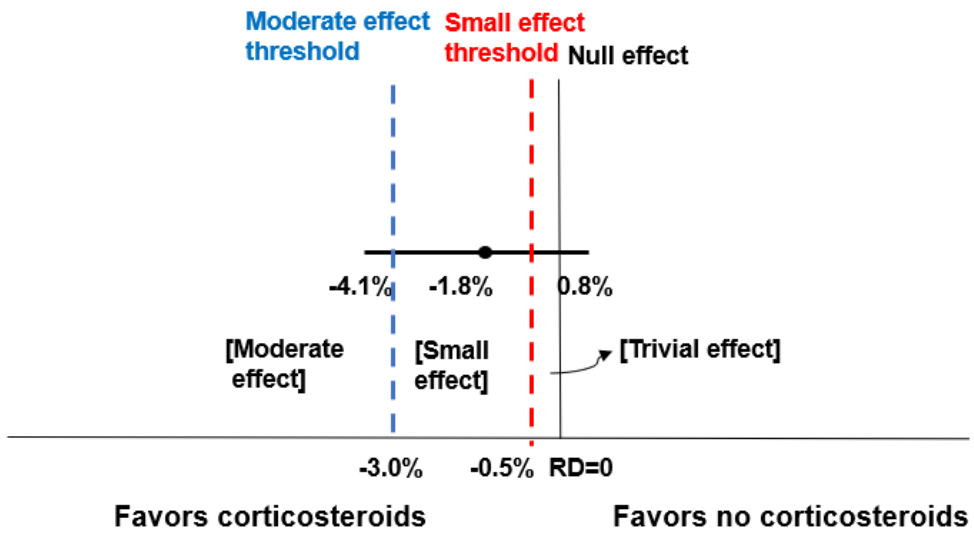


Figure 7 Application of the principles in a systematic review of corticosteroids versus placebo in patients with sepsis

Using a minimally contextualized approach, the authors could have rated their certainty that corticosteroids have an effect or have an important effect in reduction of mortality, and would have rated down for imprecision in both ratings. Using a partially contextualized approach, they could have rated their certainty that corticosteroids have a small reduction in mortality and would have rated down for imprecision.

5. Conclusion

This article has, using hypothetical and real examples, built on prior GRADE guidance regarding rating certainty of evidence and provided specific suggestions for deciding on the target of certainty rating, and how to make judgments regarding rating down for imprecision [1]. The guidance is likely to be helpful to systematic review and health technology assessment authors who should take on the challenge of being explicit regarding thresholds or ranges that underlie their ratings of certainty of evidence.

Acknowledgement

The authors would like to thank GRADE Working Group members who gave valuable comments and suggestions on the draft of this article. The authors also thank attendants of GRADE Working Group meetings who have contributed to the article during group discussions.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

BS reports receiving funding from PIPRA AG (www.pipra.ch) to conduct a systematic review and individual patient data meta-analysis on predictors of post-operative delirium in elderly for 2020-2021. BS also reports funding from Mitacs Canada, Accelerate internship in partnership with Nestlé Canada to support his graduate student stipend from 2016 to 2018. Mitacs is a national, not-for-profit organization that has designed and delivered research and training programs in Canada working with universities, companies, and both federal and provincial governments. BS also reports funding from the International Life Sciences Institute (ILSI) - North America to support his graduate work for his 2015 academic year. In 2016-2017, BS worked part-time for the Cornerstone Research Group (CRG), a contract research organization. The ILSI funding and being employed by CRG are outside the required 3 year period requested on ICJME form.

TD has received a CIHR knowledge synthesis grant and project grant for work on MIDs for PROMs.

RC reported the Parker Institute, Bispebjerg and Frederiksberg Hospital is supported by a core grant from the Oak Foundation (OCAY-18-774-OFIL).

References

1. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006
2. Noordzij M, van Diepen M, Caskey FC, et al. Relative risk versus absolute risk: one cannot be interpreted without the other. *Nephrol Dial Transplant*. 2017;32(suppl_2):ii13-ii18.

doi:10.1093/ndt/gfw465
3. Brignardello-Petersen R, Guyatt GH, Schünemann HJ, et al. Degrees of contextualization when using the GRADE approach to assess and interpret the evidence, and its relationship with network meta-analysis.(in submission)
4. Schünemann JH. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol*. 2016;75:6-15. doi:10.1016/j.jclinepi.2016.03.018
5. Brian S Alper BS, Oettgen P, Kunnamo I, et al. Defining certainty of net benefit: a GRADE concept paper. BMJ Open. 2019;9(6):e027445. doi:10.1136/bmjopen-2018-027445
6. Newcombe RG. Propagating Imprecision: Combining Confidence Intervals from Independent Sources. *Commun Stat Theory Methods*. 2011;17: 3154-3180. doi: 10.1080/03610921003764225
7. Guyatt GH, Thorlund K, Oxman AD, et al, GRADE guidelines: 13. preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol.* 2013;66(2):173-183. doi:10.1016/j.jclinepi.2012.08.001
8. Schünemann HJ, Vist GE, Higgins JPT, et al. Chapter 15: Interpreting results and drawing conclusions. Cochrane Handbook. Available at: https://training.cochrane.org/handbook/current/chapter-15#section-15-4-4-3.
9. Tikkinen KAO, Craigie S, Schünemann HJ, et al. Certainty ranges facilitated explicit and transparent judgments regarding evidence credibility. *J Clin Epidemiol*. 2018;104:46-51. doi:10.1016/j.jclinepi.2018.08.014
10. Rochwerg B., Oczkowski S.J., Siemieniuk R.A.C, et al, Corticosteroids in sepsis: an updated systematic review and meta-analysis. *Crit Care Med*. 2018;46(9):1411-1420. doi:10.1097/CCM.0000000000003262

Appendix 1 Minimally and partially contextualized approaches

1. Minimally contextualized approach

In this article, we combined what we previously called a non-contextualized approach in which reviewers use the null effect as a threshold to make judgments about the certainty of the evidence [1] with a partially contextualized approach in which reviewers use a small effect threshold, also called minimally important difference, as a threshold. We now use the label minimally contextualized approach to refer to both these approaches in which reviewers could use either the null effect or a small effect threshold.

The reasons for moving from the label of non-contextualized approach (in which no value judgments are needed) to the label of minimally contextualized approach are as follows:

First, a value judgment takes place when choosing the outcomes included in the systematic review or health technology assessment. By selecting important outcomes, those rating the certainty of evidence inevitably make a value judgment.

Second, even when rating the certainty of evidence in relation to the null effect threshold reviewers may need to make a value judgement. For example, when the point estimate is very close to the null effect, it is impossible to rate certainty of evidence that there is a trivial to null

effect without setting a threshold different than the null. To rate certainty that the true effect is trivial, reviewers need to make a value judgment and set a small effect threshold for benefit and a small effect threshold for harm. These two thresholds form a range of trivial effect (See Figure 4c in the main text).

Considering the points raised above, there is little rationale for the label non-contextualized in which no value judgment is needed.

2. Partially contextualized approach

Using a partially contextualized approach, reviewers could rate certainty that the true effect for a particular outcome, expressed in absolute terms, lies within or without the range of trivial, small, moderate or large effect [1].

In contrast to minimally contextualized approach, we are using the same conceptualization for partially contextualized approach in this article as in the previous GRADE guidance [1].

Reference

1. Hultcrantz M, Rind D, Akl EA et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006

Appendix 2 How reviewers would apply the principles to decide about the target of certainty of evidence rating when considering uncertainty from the five GRADE domains of limitation

1. Concept of certainty range

The GRADE approach for rating certainty of evidence includes five domains for rating down certainty (i.e., risk of bias, inconsistency, indirectness, imprecision, and publication bias). GRADE uses the concept of certainty range to characterize uncertainty that considers all the five domains of limitations [1,2]. The uncertainty associated with one of these domains of limitations, imprecision, can be quantified by examining confidence intervals (CI) (or credible intervals for Bayesian analysis) [1,2]. The extent of uncertainty associated with the other four domains of limitations is not, thus far, amenable to quantification. Therefore, we still do not know how to quantify the certainty range.

2. Application of the principles when considering the certainty range

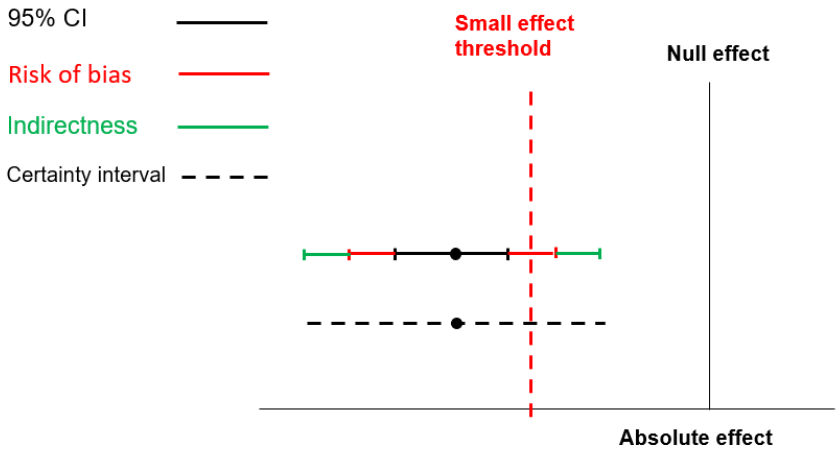
Conceptually, the other four GRADE domains extend and modify the distribution function of uncertainty for the best estimate of effect beyond that defined by the 95%CI [2]. The width of the certainty range would depend on the extent of concerns regarding the other four domains: the greater the concerns, the less is known about the width and shape of the probability distribution of the estimates in that range.

For example, consider a situation in which reviewers have serious concerns regarding risk of bias and indirectness, but have no reason to believe that the 95% CI would widen more on one side than the other (in other words, both risk of bias and indirectness can be acting in both directions). As presented in Appendix Figure 1a, the point estimate would not change, and the certainty range would widen beyond the 95% CI due to risk of bias and indirectness. Using a

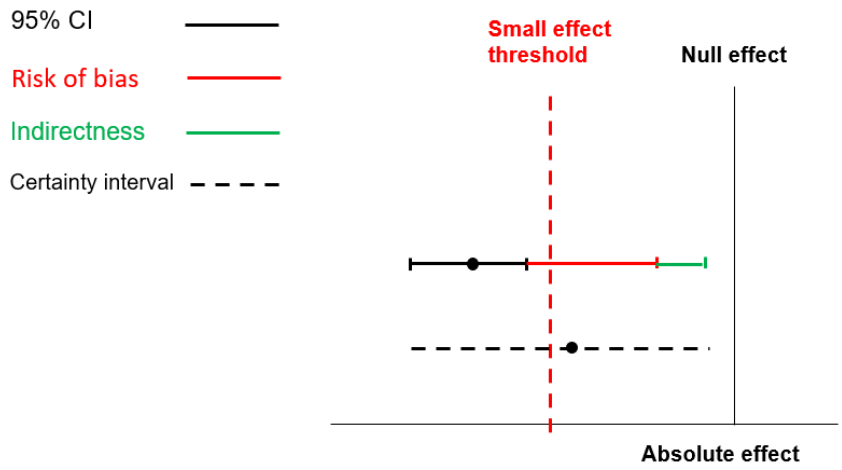
minimally contextualized approach, because the best estimate still suggests an effect greater than the small effect threshold, reviewers would rate certainty that the true effect is small, and would rate down certainty for risk of bias and indirectness but not for imprecision.

In another situation, if the risk of bias and the indirectness have a clear direction, the certainty range would widen in only one direction and the best estimate which presents the most probable true effect might move from one side of the threshold to another. As presented in Appendix Figure 1b, if reviewers are confident that risk of bias and indirectness overestimated the treatment effect and the best estimate should be moved from the left side of the threshold to the right side, they would rate certainty that the true effect is a trivial effect, and would rate down their certainty of evidence for risk of bias and indirectness but not for imprecision. Situations in which reviewers are aware of the direction of bias, and have a clear enough sense of its magnitude to confidently move the point estimate, are currently few and far between.

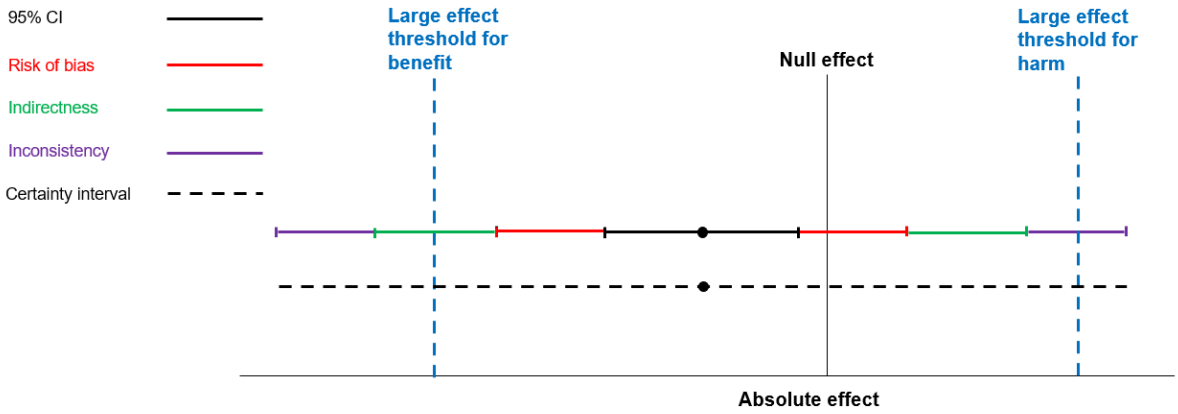
Conceptually, if reviewers could quantify the certainty range, when the uncertainty associated with some or all of the five domains of limitation is great that the certainty range becomes extremely wide, reviewers could abandon being explicit about the target of certainty rating (Appendix Figure 1c).



(a) Reviewers are unaware of the direction to which the concerns of the other four domains of limitation would widen the certainty range than the 95% CI



(b) Reviewers are aware of the direction to which the concerns of the other four domains of limitation would widen the certainty range than the 95% CI



(c) The certainty range is extremely wide, reviewers could abandon being explicit about the target of certainty rating

Appendix 2 Figure 1 Determine the target of certainty rating when consider the certainty range

References

1. Tikkinen KAO, Craigie S, Schünemann HJ, et al. Certainty ranges facilitated explicit and transparent judgments regarding evidence credibility. *J Clin Epidemiol*. 2018;104:46-51. doi:10.1016/j.jclinepi.2018.08.014
2. Schünemann JH. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol*. 2016;75:6-15. doi:10.1016/j.jclinepi.2016.03.018

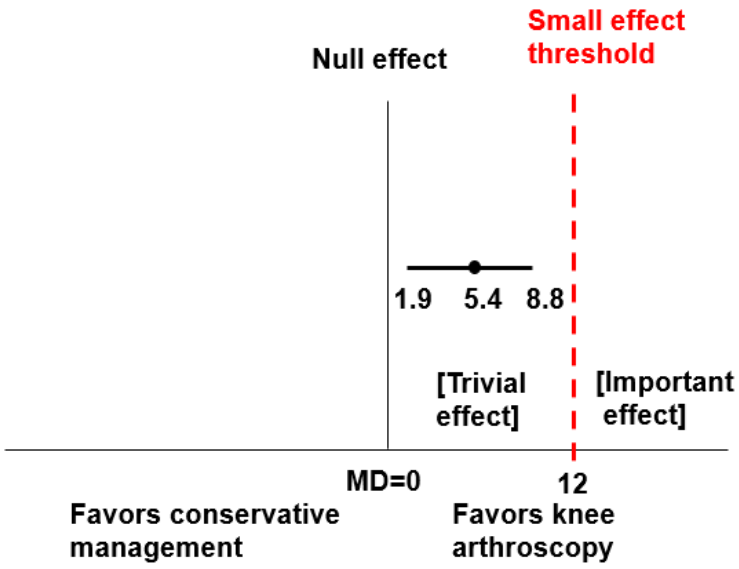
Appendix 3 Application of principles in real examples

Here we present additional examples from published systematic reviews to illustrate how reviewers could use the principles in determining the target of certainty of evidence rating. The authors of these reviews did not always follow the guidance we suggest here – thus illustrating the desirability of the clarifications we present in this article.

Example 1: Determining the target of certainty of evidence rating for continuous outcome

A systematic review of knee arthroscopy versus conservative management including 1,231 patients with degenerative knee disease from 10 randomized controlled trials (RCTs) yields a pooled mean difference in change of pain score from baseline until 3 month of 5.4 points higher on a 100-point scale (the higher score the better), with a 95% confidence interval (CI) from 1.9 to 8.8 points higher (Appendix 3 Figure 1) [1]. The authors could have used a minimally contextualized approach and rated their certainty in relation to the null effect, in which case there would have been no need to rate down for imprecision.

The authors, however, chose to rate certainty in relation to what they defined - based on a systematic review of minimally important difference (MIDs) offered for the relevant pain instrument - as an MID of 12 points (i.e. small effect threshold) [2]. Considering the point estimate, the authors rated their certainty that the true effect is smaller than the MID (i.e. the true effect is trivial). Further, because the 95% CI did not cross the threshold of 12, they did not rate down the certainty for imprecision. The systematic review authors found no other limitations in the body of evidence, and thus concluded with high certainty that knee arthroscopy does not result in an important reduction in pain when compared with conservative management in patients with degenerative knee disease [1].



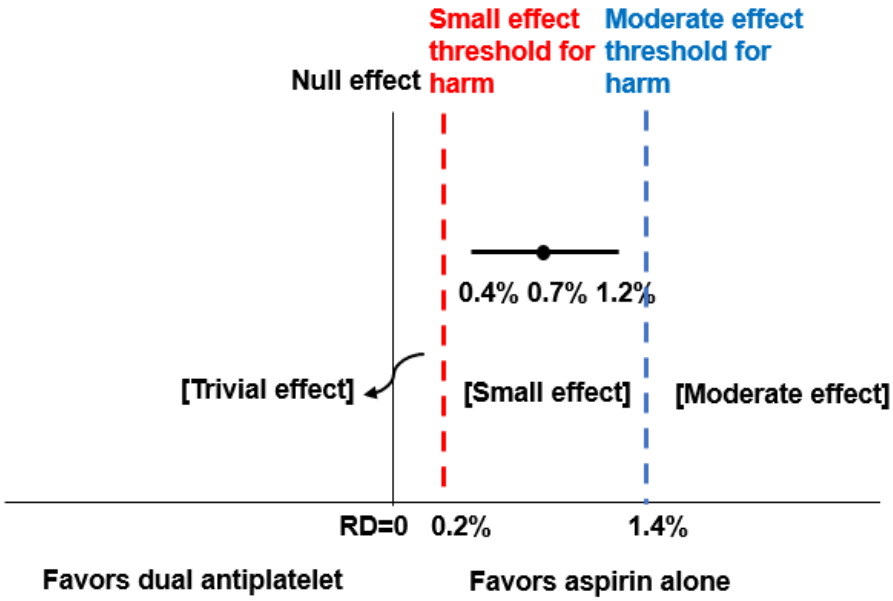
Appendix 3 Figure 1 Determining the target of certainty rating for a continuous outcome (Example 1)

Example 2: Determining the target of certainty of evidence rating when effect estimates suggest an increase in harmful outcome

Consider a systematic review of clopidogrel and aspirin (dual antiplatelet therapy) versus aspirin alone in patients with minor ischaemic strokes or at high-risk of transient ischaemic attacks [3]. The point estimate of the risk difference (RD) for extracranial bleeding from the pooled analysis of

10,075 patients in three studies is 0.7 more bleedings per 100 patients with a 95% CI from 0.4 more to 1.2 more bleedings per 100 patients in dual antiplatelet therapy compared with clopidogrel alone (Appendix 3 Figure 2) [3].

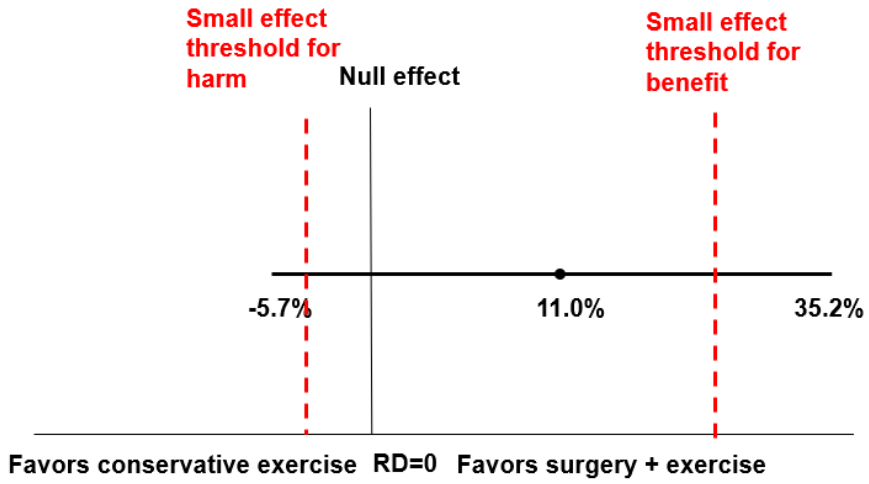
Using a partially contextualized approach, had the authors set the small effect threshold for harm at 0.2 more bleedings per 100 patients and the moderate effect threshold for harm at 1.4 more bleedings per 100 patients, the authors could then have rated their certainty that dual antiplatelet therapy results in a small increase in extracranial bleeding, compared with aspirin alone. Because the 95% CI does not overlap either with the small or the moderate harm thresholds, they would not rate down the certainty due to imprecision. As there were no other limitations of the body of evidence, the review team would have high certainty that the harm is small.



Appendix 3 Figure 2 Determining the target of certainty rating when results suggest an increase in harmful outcome (Example 2)

Example 3: When the 95% CI is extremely wide, reviewers could abandon being explicit about the target of certainty of evidence rating

Consider a systematic review of arthroscopic surgery plus postoperative exercise therapy versus conservative exercise therapy in adults with subacromial pain syndrome (SAPS) [4]. The authors found one RCT that addressed this issue. The RCT revealed that surgery plus exercise yielded 11 more patients achieving success per 100 patients (defined as "no shoulder problems at all"/ "healed completely" or "much better"), with a 95% CI from 5.7 fewer to 35.2 more per 100 patients at 6 months compared with conservative exercise (Appendix 3 Figure 3). One might well judge that the 95% CI crosses both the threshold of large benefit and the threshold of small but important harm. Considering the width of the 95% CI, the systematic review team were very uncertain whether surgery improves or worsens global perceived effect at 6 months [4]. In such situations, there would be no need for authors to decide about the target for certainty rating, and they would rate down two levels for imprecision.



Appendix 3 Figure 3 When the 95% CI is extremely wide, reviewers could abandon being explicit about the target of certainty rating (Example3)

References

1. Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al, Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open.* 2017;7(5):e016114. doi:10.1136/bmjopen-2017-016114 2017;7:e016114.
2. Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations. *BMJ Open*. 2017;7(5):e015587. doi:10.1136/bmjopen-2016-015587
3. Hao QK, Tampi M, O'Donnell M, et al. Clopidogrel plus aspirin versus aspirin alone for acute minor ischaemic stroke or high risk transient ischaemic attack: systematic review and meta-analysis. *BMJ*. 2018;363:k5108.. doi:10.1136/bmj.k5108
4. Lähdeoja T, Karjalainen T, Jokihaara J, *et* al. Subacromial decompression surgery for adults with shoulder pain: a systematic review with meta-analysis. *Br J Sports Med*. 2020;54(11):665-673. doi:10.1136/bjsports-2018-100486

Chapter 4

Rating certainty when the target threshold is the null, and the point estimate is close to the null

(Unpublished, Zeng L will be the first author)

Abstract

Objective: To identify the challenges faced by review authors when deciding on the target of certainty rating and making imprecision judgments when the point estimates is close to an initial threshold of interest, and to propose strategies to overcome those challenges.

Study Design and Setting: Based on GRADE guidance addressing decisions regarding the target of certainty of evidence rating, a project group within the GRADE Working Group (GWG) conducted iterative discussions and presentations at GWG meetings.

Results: When the point estimate falls close to a threshold (i.e., the null effect) indicating an unimportant effect, rating certainty in a non-zero effect is misleading. When the confidence interval is narrow rating down for imprecision (because the confidence interval crosses the null) is counterintuitive. Facing these two challenges, review authors may judge whether the point estimate i) is clearly greater than an minimally important difference (MID); if so they can still rate certainty in a non-zero effect, ii) is well below the MID; they will then switch to rate certainty in a little to no effect, or iii) falls in the range of uncertainty around the MID (i.e., a range of plausible MIDs), in which case they can either switch or not switch the target of certainty rating. After switching the target of certainty rating from a non-zero effect to a little to no effect, making the imprecision judgment requires deciding whether the confidence interval overlaps with the MID for benefit or the MID for harm (that form the range of little to no effect). If so they will rate down at least one level for imprecision.

Conclusion: Review authors should note the challenges in deciding on the target of certainty rating when the point estimates is close to an initial threshold of interest. Based on the relation of the point estimate and the MID review authors can make optimal choice of the target of certainty rating and make imprecision judgments accordingly.

1.Introduction: GRADE guidance thus far

In 2017, GRADE published a key paper in which the authors clarified what it is in which those using GRADE are rating their certainty – the target of the certainty rating [1]. Previous to that paper, GRADE had specified the target as certainty in the point estimate. The 2017 paper pointed out the flaws in this conceptualization and suggested an alternative that has since become core GRADE guidance: we are rating our certainty that the true effect lies on one side of a threshold or in a particular range.

Possible thresholds include the null (the focus of the current paper) and the minimally important difference (MID) (also called the small effect threshold). Ranges can include trivial, small, moderate and large effects [1]. One can also rate certainty in net benefit. In a paper published in 2021, GRADE offered additional guidance on deciding the target of the certainty rating, clarifying that the key determinant of the choice is where the point estimate lies in relation to the chosen threshold [2].

Following the guidance in these two papers, systematic review authors – the target audience for this guidance - must, when rating certainty of evidence, always identify the chosen threshold or range, and rate certainty accordingly. If they believe it is most relevant to their audience, and if in addition review authors prefer to avoid making the value judgements that are required when they choose the MID or some other magnitude of effect as the threshold, review authors will choose the null effect as the threshold of interest. This paper deals with possible complications that may arise when one has chosen the null as the target, and the point estimate is close to the null. Our previous paper briefly described the core guidance for this situation[2]; the current paper presents much more detail and informative examples, and deals with uncertainties that arise when following the prior guidance.

2. The choice of the null as the target

Having chosen the null as the threshold, for point estimate A in Figure 1 that suggests an appreciable benefit, review authors would without ambivalence rate their certainty in a non-zero effect (i.e., the target of certainty rating). For point estimate B in Figure 1, as the point estimate is very close to the null, according to current GRADE guidance[2] review authors could consider either rating certainty in relation to the initial threshold of interest (i.e., the null effect) or switching to rate certainty of evidence to two adjacent thresholds (i.e., the MID for benefit and the MID for harm). In this paper, we will describe in some detail how reviewer should handle this choice, beginning, in the next section, with illustration of the problem.

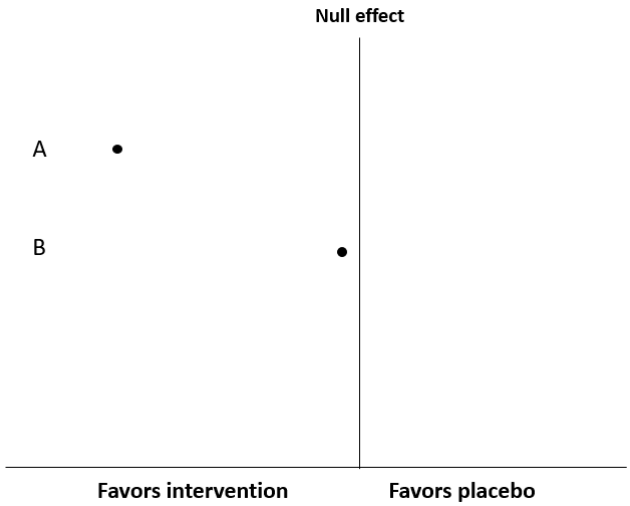


Figure 1 Target of certainty ratings in relation to the null effect

Figure legend: For point estimate A, review authors could rate their certainty in a non-zero effect, while for point estimate B, as the point estimate is very close to the null, challenges might arise if the authors continue to rate certainty in relation to the initial threshold.

3. Two major challenges in certainty rating when the point estimate is close to the null and the confidence interval crosses the null

3.1 When the confidence interval is narrow rating down for imprecision is counterintuitive

Consider a hypothetical systematic review of a new intervention X for patients at risk of myocardial infarction. Review authors wish to avoid making the value judgements required for using the MID as a threshold, and therefore choose to rate certainty in relation to the null. For situation A in Figure 2, as the point estimate favors intervention X, they would rate their certainty that intervention X reduces the risk of myocardial infarction (i.e., a non-zero effect exists) and would, because the confidence interval does not overlap the null, not rate down for imprecision. For situation B, the review authors would still rate their certainty that the intervention reduces the risk of myocardial infarction, but because in this situation the confidence interval overlaps the threshold of interest, rate down for imprecision.

For situation C, the confidence interval overlaps the null and therefore, if the authors continue to rate certainty in a non-zero effect (i.e., rating certainty in relation to the null), because the confidence interval crosses their threshold, they would rate down for imprecision. This choice would, however, be both counterintuitive and inappropriate: the very narrow confidence interval precludes rating down for imprecision.

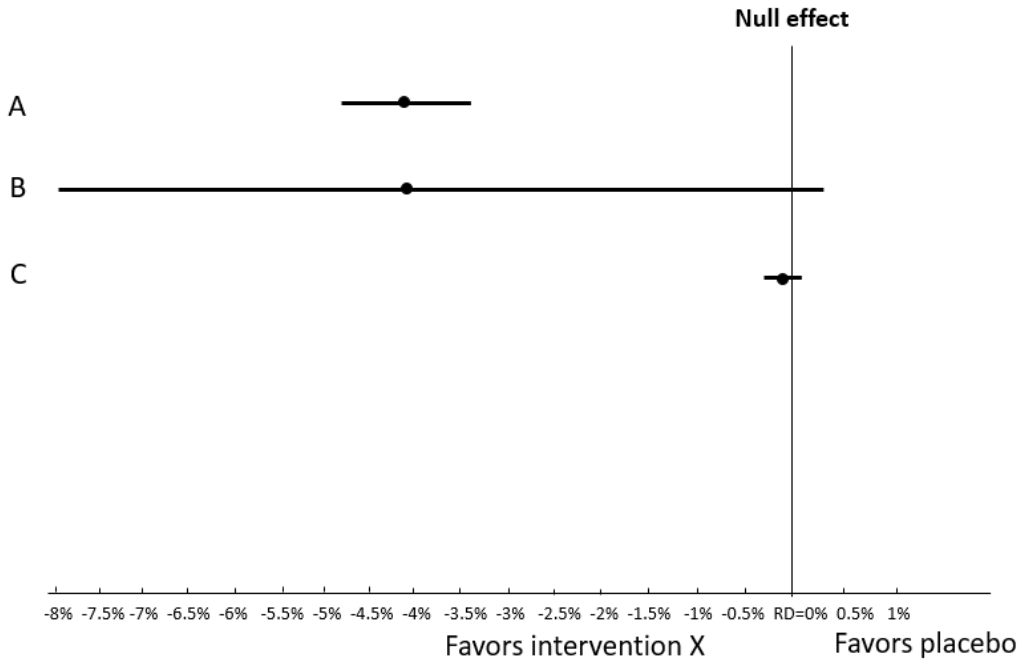


Figure 2. A hypothetical systematic review of intervention X on myocardial infarction in patients with risk of myocardial infarction

Figure legend: For situation A, review authors could rate their certainty that intervention X reduces the risk of myocardial infarction and would not rate down for imprecision. For situation B, they could still rate their certainty in a non-zero effect and rate down for imprecision. For situation C, if the review authors choose to rate certainty in a non-zero effect, it would be counterintuitive to rate down for imprecision.

3.2 When the point estimate indicates an unimportant effect rating certainty in a non-zero effect is misleading

In reflecting on situation C in Figure 2, reviewers also face a second source of discomfort. This discomfort arises from rating certainty in the benefit of the intervention when the point estimate, close to the null, in fact suggests that there is little to no effect of intervention.

Considering both issues (avoiding rating down for imprecision when the confidence interval is very precise, and rating certainty in a benefit when the point estimate suggests little to no benefit), reviewers need specific guidance for addressing how to choose the target of the certainty rating they initially plan to rate in a non-zero effect and the point estimate turns out to be close to the null. This paper, using detailed examples, provides this guidance, illustrating how review authors can achieve explicitness regarding their decisions when the point estimate is close to the null.

4. Methods

A steering group drafted an initial version of this paper that focused on situations in which review authors choose the null as the threshold. In November, 2022 we first presented the paper online in a small group discussion at the GRADE Working Group (GWG) meeting. After iterative discussions within the project team, we presented the revised paper in a webinar within the Certainty in Evidence Project Group in April, 2023 and in large group discussions at Split, London and Glasgow

in May and September, 2023. Following these discussions we made substantive changes based on feedback from the entire GWG. At the time of thesis submission, this paper is not approved as a GRADE concept paper.

5. The solution to the challenges: switching the target of certainty rating from a non-zero effect to a little to no effect

For situation C in Figure 2, recognizing that the null may indeed represent the true treatment effect – or if not the null, at least a value very near the null that represents a clearly unimportant effect – offers the solution to the two problems. Reviewers can switch from rating certainty in a non-zero effect (choosing the null as the threshold of interest) to rating certainty in little to no effect (choosing a range of little of no effect as the range of interest). In the remainder of this article, we will address issues that arise in making this switch in the target of the certainty rating.

Consider a real example: a systematic review of lower blood pressure target (≤ 135/85 mmHg) versus standard blood pressure target (≤ 140–160/90–100 mmHg) for patients diagnosed with cardiovascular disease and with high blood pressure [3]. A meta-analysis of seven randomized controlled trials (RCTs) including 9,595 patients reported a point estimate of 0 fewer death per 100 patients, with a confidence interval from 10 fewer to 10 more death per 1000 patients (Figure 3).

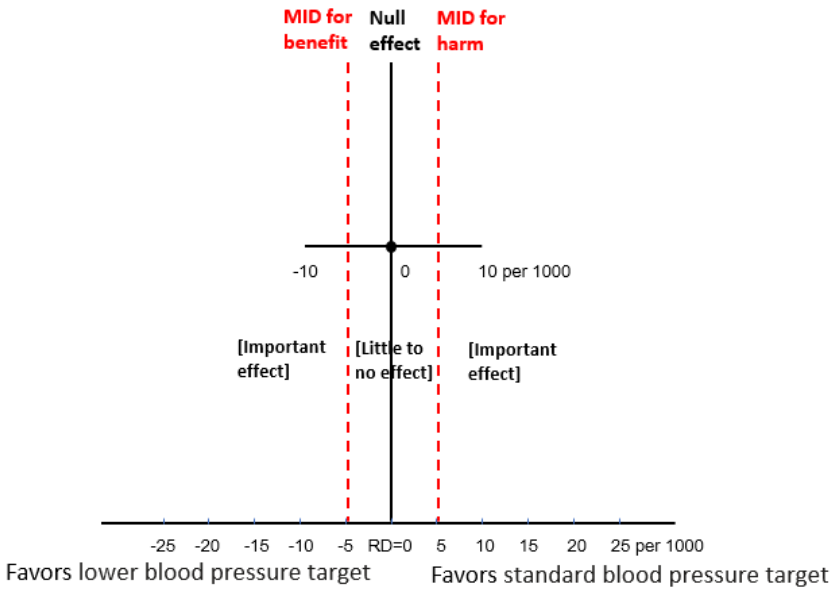


Figure 3 A meta-analysis of lower versus standard blood pressure target on total mortality in patients with cardiovascular disease and high blood pressure

While the review authors did not specify their target of certainty rating, we will consider that they started with the intention of rating certainty in a non-zero effect. As the point estimate falls very close to the null, rating certainty in either benefit or harm becomes counter-intuitive and likely misleading, and changing the target of the certainty rating to certainty in a little to no effect would be advisable.

6. When to switch the target of certainty of evidence rating from a non-zero effect to a little to no effect

In situations when the confidence interval crosses the null, review authors might ask how close the point estimate needs to be to the null before they should consider switching the target of certainty rating from a non-zero effect to little to no effect. The previous GRADE guidance does not provide an answer. We suggest that, once review authors, considering the point estimate, conclude that it evidently represents an unimportant effect (i.e., when the point estimate is less than the MID), they should consider switching the target of certainty rating. This highlights the necessity of, as soon as reviewers seriously consider switching, they need to confront an issue they were hoping to avoid: specifying the MID.

Having specified the MID, reviewers will then determine the relation between the point estimate and the MID: point estimate greater than the MID, continue to rate certainty in a non-zero effect, point estimate less than the MID, switch to rating certainty in little or no effect. Reviewers may, however, experience considerable discomfort in applying this guidance to situation B in Figure 4 (i.e., the point estimate is very close to the MID). They might reasonably consider whether it remains preferable to rate certainty in a non-zero effect.

The source of this discomfort may be awareness of the arbitrariness, or at least the uncertainty, associated with estimating an MID. As discussed above, setting the MID always involves a value judgement, and making such a judgement is often challenging. Given the usual paucity of evidence regarding patients' values, and the variation in patients' values, uncertainty regarding the MID is invariably appropriate. Thus, review authors may consider a range of plausible MIDs from the largest to the smallest plausible MID (Figure 4).

If the point estimate falls above the value that review authors view as representing the largest plausible MID (A in Figure 4), they would certainly rate certainty in a non-zero effect. If, on the other hand, the point estimate falls below the value one has designated as the smallest plausible MID, and the point estimate therefore represents an unimportant effect (C in Figure 4), they would confidently switch the target of certainty rating to a little to no effect.

If review authors consider that the point estimate falls in the range of uncertainty regarding the MID (that is, authors are uncertain if the point estimate is above or below some true MID), particularly if the point estimate falls below their best estimate of the MID (B in Figure 4), they may or may not switch the target of certainty rating to a little to no effect. Either option would be reasonable.

For systematic reviewers, specifying the best estimate of MID or its plausible range might be challenging and uncomfortable. It turns out it may also be unnecessary. All review authors need to do is to judge whether their specific point estimate is *i)* clearly greater than the MID (they can then still rate certainty in a non-zero effect), *ii)* well below the MID (they will then switch to rate certainty in a little to no effect), or *iii)* the point estimate falls in the range of uncertainty around the MID (in which case they can either switch or not switch the target of certainty rating). They can make this judgement without specifying an exact MID, or an exact range of certainty around the MID. Even if

they choose to avoid specification of an exact MID and an exact plausible range, understanding the logic underlying the choice of switching or not switching will be helpful.

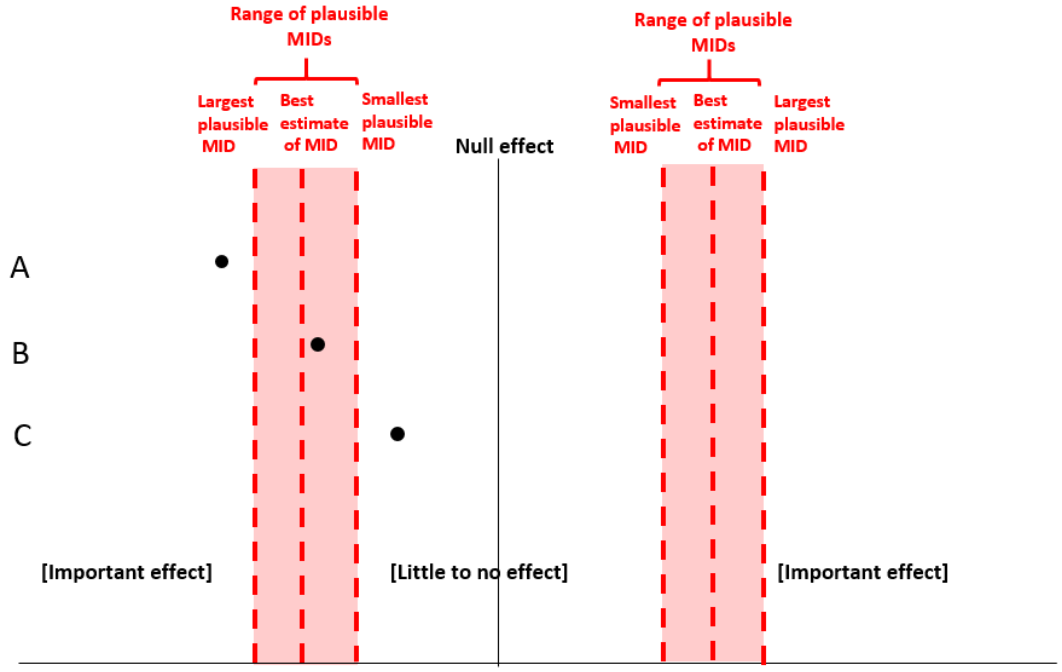


Figure 4 Plausible range of MIDs

Figure legend: Situation A represents a situation in which review authors are confident that the point estimate is above the MID (indicating an important effect). Situation C represents a situation in which review authors are confident that the point estimate is below the MID (indicating an unimportant effect). Situation B represents a situation in which review authors are uncertain whether the point estimate is above or below the MID.

7. Imprecision rating after switching the target of certainty rating

After switching the target of certainty rating, making the imprecision judgment requires deciding whether the confidence interval overlaps either of the two thresholds (i.e., the MID for benefit and MID for harm) that form the range of little to no effect. That is, when rating certainty in a little to no effect, reviewers judge whether the confidence interval indicates that the true effect might still be an important benefit or an important harm.

As discussed above, the decision regarding the MID is likely to involve at least some uncertainty or arbitrariness. Although a detailed discussion of how to choose MIDs is outside the scope of this paper, we will briefly discuss the issue.

Formal approaches to help establish MIDs are now available [4-7]. Alternatively, searching the literature for studies of values and preferences, and health state utilities, may be helpful. If the outcome is a patient-reported measurement instrument, authors are likely to find a relevant literature establishing a suggested MID [8]. Review authors can acknowledge the uncertainty or arbitrariness and potentially comment on how reasonable alternative choices would impact on decisions regarding rating down for imprecision.

In the blood pressure example, had review authors set the MID of mortality as a difference around 5 per 1000 (Figure 3), as the confidence interval overlaps with both boundaries of the range (i.e., the confidence interval includes both important benefit and important harm), they would certainly rate down at least once for imprecision and may rate down twice [9]. If the authors had no concerns regarding the other four GRADE domains, the plain language summary would be that the lower blood pressure target likely has (rating down once for imprecision) or may have (rating down twice for imprecision) little to no effect on mortality.

Without explicitly specifying an MID threshold, the review authors can also make judgement on imprecision: they can consider whether the boundary of the confidence interval clearly falls below the MID (if so, not rate down for imprecision), or clearly falls above the MID (if so, rate down for at least one level for imprecision), or falls within the range of plausible MIDs (if so, they cannot avoid considering the best estimate of the MID; and if the point estimate falls below the best estimate of the MID, not rate down for imprecision, otherwise, rate down at least one for imprecision).

8. Examples of deciding on the target of certainty rating when the point estimate is close to the null and the confidence interval crosses the null

Consider a systematic review addressing corticosteroids versus no corticosteroids in patients with sepsis [10]. A meta-analysis of 17 RCTs with 4243 participants reported a point estimate of 3 more gastrointestinal bleeds per 1000 patients in patients randomized to corticosteroids, with a confidence interval from 5 fewer to 13 more per 1000 patients (Figure 5).

Had review authors initially planned to rate certainty in a non-zero effect, as the point estimate turned out to be close to the null effect, they would start to consider switching the target of the certainty rating. Without considering what the exact value of the MID might be, they might reasonably conclude that an increase of 3 in 1,000 clearly represents an unimportant effect (i.e., clearly falls below the MID). Having made this inference, they would switch the certainty target to little to no effect and move to consider the boundaries of the confidence interval. If their judgement is that 13 per 1000 increase also represents an unimportant effect, as does a 5 per 1000 decrease, they would not rate down for imprecision, and can do so without deciding at what value above 13 per 1000 the MID lies. Assuming the authors have no concerns for the remaining four GRADE domains, the plain language summary would state that corticosteroids have little to no effect on gastrointestinal bleeding.

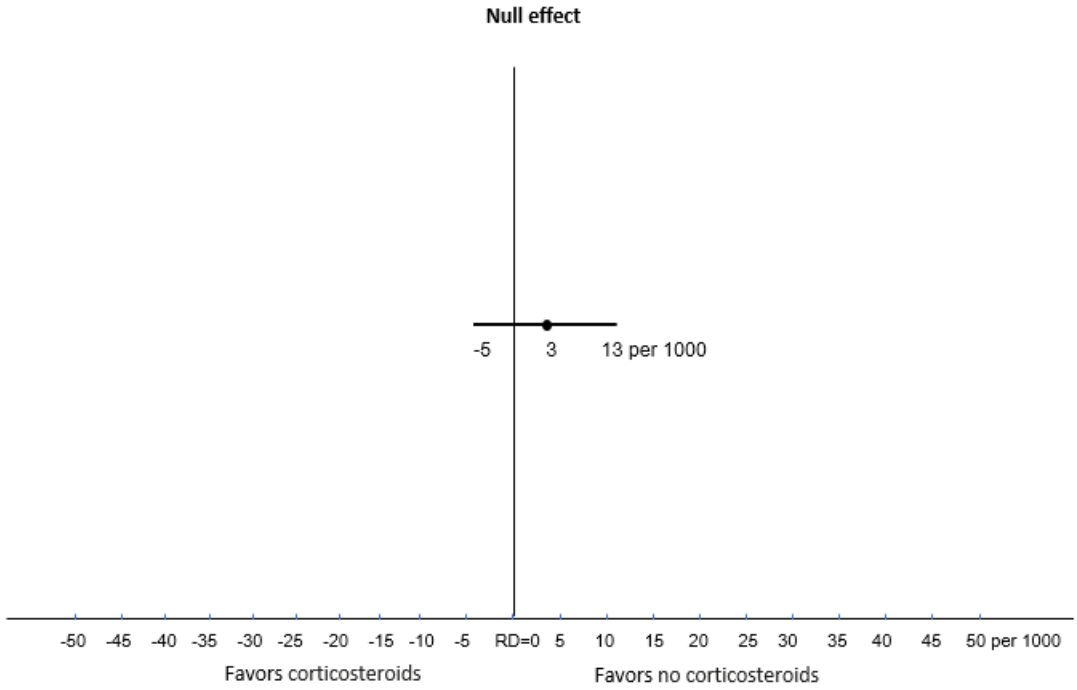


Figure 5 A meta-analysis of corticosteroids versus no corticosteroids on gastrointestinal bleeding in patients with sepsis

Consider another systematic review of phenobarbital versus placebo or no intervention in preterm infants with or at risk of intraventricular haemorrhage (IVH) [11]. A meta-analysis of 10 RCTs with 792 patients reported a point estimate of 20 fewer IVHs per 1000 patients, with a confidence interval from 70 fewer to 30 more per 1000 (Figure 6).

Without specifying the explicit boundaries of the range of plausible MIDs, the authors could have considered the point estimate of a reduction of 20 per 1000 patients lies within that range (but uncertain whether it is above or below the best estimate of MID). They could still rate certainty in a non-zero effect. If they did so, as the confidence interval simultaneously includes a 30 per 1000 increase, they would rate down one level for imprecision (if they consider 30 per 1000 increasing unimportant and 70 per 1000 decreasing important) or two levels for imprecision (if they consider 30 per 1000 increasing important). Alternatively, they could switch to rate certainty in a little to no effect and rate down one level for imprecision (if they consider 30 per 1000 increasing still unimportant and 70 per 1000 decreasing important) or two levels (if 30 per 1000 increasing important).

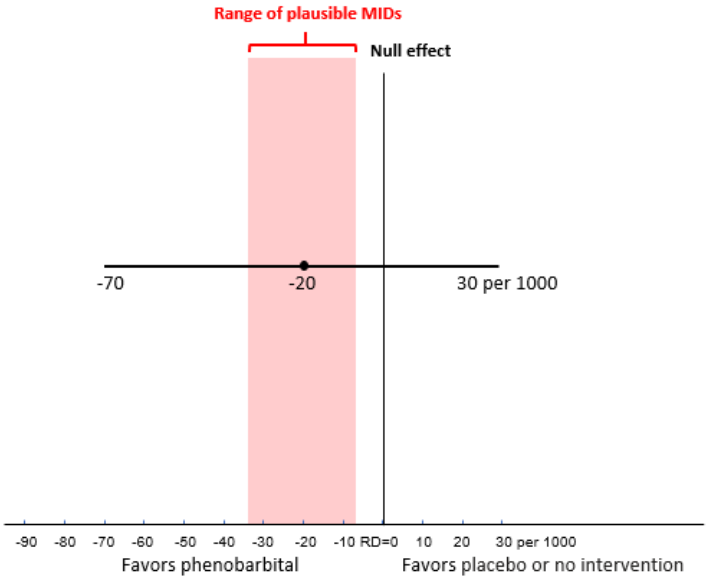


Figure 6 A meta-analysis of phenobarbital versus placebo on intraventricular haemorrhage (IVH) in infants with preterm infants with or at risk of IVH

This approach for considering the target of certainty rating also applies to continuous outcomes. Consider a systematic review of closed versus open kinetic chain exercises for patients with patellofemoral pain syndrome [12]. A meta-analysis of three RCTs including 122 patients reported a point estimate of an increase of 0.03 points on a 0 to 10 Visual Analogue Scale (VAS) with a confidence interval from a decrease of 0.37 to an increase of 0.76 points (Figure 7). The empirical evidence suggests that the MIDs on a 0-10 VAS range from one to two points [13, 14].

Had the review authors initially planned to rate certainty in a non-zero effect, as the point estimate falls below the range of plausible MIDs and the confidence interval crosses the null, they would switch the target of certainty rating from null effect to little to no effect. As the entire confidence interval falls within the range of little to no effect, the authors would not rate down for imprecision. Given the review authors have no concerns on the remaining four GRADE domains, the plain language summary would be closed kinetic chain exercises, compared with open kinetic chain exercises, have little to no effect on pain intensity.

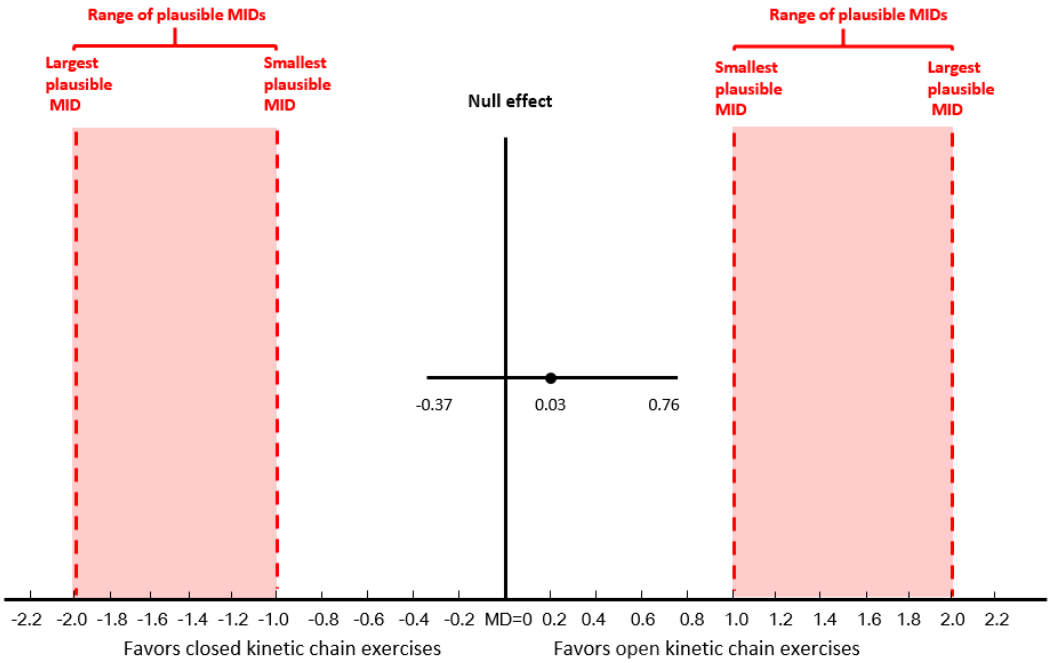


Figure 7 A meta-analysis of open versus closed kinetic chain exercises on pain intensity in patients with patellofemoral pain syndrome

9. A very small effect with a confidence interval not crossing the null

Another potential problematic situation that we have not addressed so far occurs when the point estimate is near the null but a precise confidence interval does not cross the null (as A in Figure 8). Such a point estimate may represent an unimportant effect. One might ask the question: does a definitive conclusion that a non-zero effect (at least with respect to imprecision) exists, give a misleading impression that the effect is important?

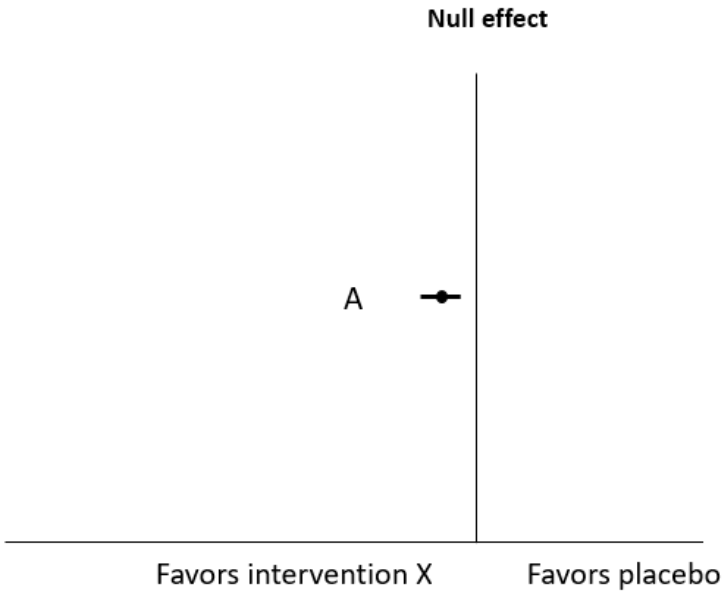


Figure 8 Situation in which the point estimate is close to the null and the confidence interval does not cross the null effect

In choosing the null as the certainty rating, review authors consciously avoid decisions regarding the MID. In doing so, they have left the decision regarding the importance of the effect to those who, in the future, will make decisions regarding use of the intervention of interest. Authors, concerned about giving readers the wrong message, may note that the small non-zero effect their results indicate may be unimportant. If they feel very uncomfortable about a misleading message, they could change their decision and rate certainty in little or no effect, conclude little or no effect.

Consider a systematic review of subacromial decompression surgery in addition to exercise therapy versus exercise therapy alone for patients with shoulder pain. A meta-analysis of four RCTs including 399 patients reported a point estimate of a decrease of 1.0 points on a 0 to 10 VAS at one year with a confidence interval from a decrease of 1.6 to 0.4 points (Figure 9) [15]. Unlike the previous example (in which we assume the review authors identified a range of plausible MIDs), the review authors explicitly defined an MID as 1.5 points [16]. In such case (i.e., the point estimates falls below the MID and the entire confidence interval falls on the side that favors subacromial decompression surgery), the review authors could still rate certainty that the subacromial decompression surgery decreases pain intensity (i.e., rating certainty in a non-zero effect) and would not rate down for imprecision. Indeed, if there were no problems in any other domain they would conclude high certainty evidence. As the point estimate clearly indicates an unimportant effect, it might well be advisable to alert their readers that the small non-zero effect may be trivial. Were they to judge this warning as inadequate, they could decide to switch to rate certainty in little or no effect – in which case they would rate down for imprecision because the confidence interval crosses their importance threshold.

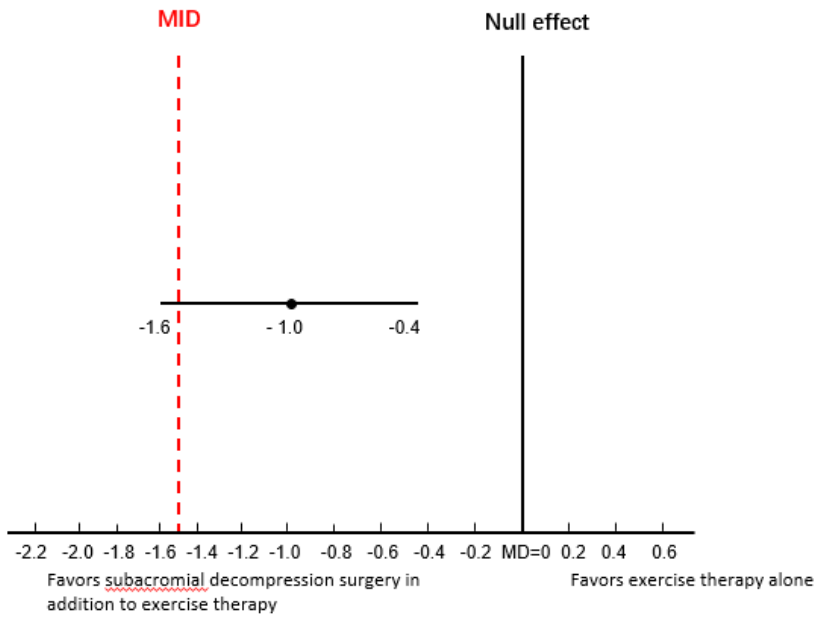


Figure 9 A meta-analysis of subacromial decompression surgery in addition to exercise therapy versus exercise therapy alone on pain intensity in patients with shoulder pain

10. Conclusion

This article has, using hypothetical and actual examples, introduced the problems that arise when review authors initially decide to rate certainty in a non-zero effect, and the point estimate falls close to the null. Our suggested approach considers whether the point estimate is possibly or certainly less than the MID: if certainly so (i.e., falls below the smallest plausible MID), GRADE suggests switching to rate certainty in a little to no effect. If possibly so (i.e., falls within the range of plausible MIDs, particularly below the best estimate of MID), options of continuing to rate certainty in a non-zero effect or switching remain.

Review authors intending to switch the target of the certainty rating if they find a point estimate sufficiently near the null will ideally pre-specify this plan in their protocol. They could further specify that in doing so, they would consider if the point estimate is clearly less than the MID – if so, they would proceed with the switch. If, on the other hand, the point estimate is in a range that may or may not be less than the MID (the range of the plausible MID) they could specify either that they will retain the original target (the null) or switch to rating certainty in little to no effect. Pre-specifying how they will use our suggested approach to the problem of a point estimate near the null will usefully clarify their thinking before they begin their review.

Alternatively, for deciding the target of certainty rating, the review authors do not need to consider the MID (or plausible range of MIDs) unless the point estimate turns out to be close to the null (i.e., they seriously consider switching the target of certainty rating from a non-zero effect to a little to no effect).

References

1. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006
2. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *J Clin Epidemiol.* 2021;137:163-75. doi: 10.1016/j.jclinepi.2021.03.026
3. Saiz LC, Gorricho J, Garjón J, et al. Blood pressure targets for the treatment of people with hypertension and cardiovascular disease. *Cochrane Database Syst Rev.* 2022;11(11):CD010315. doi:10.1002/14651858.CD010315.pub5
4. Morgano GP, Mbuagbaw L, Santesso N, et al. Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomised methodological study (GRADE-THRESHOLD). *BMJ Open*. 2022;12(3):e053246. doi:10.1136/bmjopen-2021-053246
5. Zeng L, Li S-A, Helsingen LM, et al. A novel framework for incorporating patient values and preferences in making guideline recommendations: the presentation and evaluation of guideline panel surveys. *J Clin Epidemiol*. 2023;S0895-4356(23)00175-0. doi:10.1016/j.jclinepi.2023.07.003
6. Helsingen LM, Zeng L, Siemieniuk RA, et al. Establishing thresholds for important benefits considering the harms of screening interventions. *BMJ Open.* 2020;10(12):e037854. doi: 10.1136/bmjopen-2020-037854

7. Neumann I, Quiñelen E, Nahuelhual P, et al. Using Explicit Thresholds were valuable for judging Benefits and Harms in partially contextualized GRADE Guidelines. *J Clin Epidemiol*. 2022;147:69-75. doi:10.1016/j.jclinepi.2022.03.017
8. Devji T, Carrasco-Labra A, Guyatt G. Mind the methods of determining minimal important differences: three critical issues to consider. *Evid Based Ment Health*. 2021;24(2):77-81. doi:10.1136/ebmental-2020-300164
9. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol*. 2022;150:216-224. doi:10.1016/j.jclinepi.2022.07.014
10. Rochwerg B, Oczkowski SJ, Siemieniuk RAC, et al. Corticosteroids in sepsis: an updated systematic review and meta-analysis. *Crit Care Med*. 2018;46(9):1411-1420. doi:10.1097/CCM.0000000000003262
11. Romantsik O, Smit E, Odd DE, Bruschettini M. Postnatal phenobarbital for the prevention of intraventricular haemorrhage in preterm infants. *Cochrane Database Syst Rev*. 2023;3(3):CD001691. doi:10.1002/14651858.CD001691.pub4
12. van der Heijden RA, Lankhorst NE, van Linschoten R, et al. Exercise for treating patellofemoral pain syndrome. *Cochrane Database Syst Rev*. 2015;1:CD010387. doi:10.1002/14651858.CD010387.pub2
13. Kelly A-M. The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. *Emerg Med J*. 2001;18(3):205-207. doi:10.1136/emj.18.3.205
14. Olsen MF, Bjerre E, Hansen MD, et al. Pain relief that matters to patients: systematic review of empirical studies assessing the minimum clinically important difference in acute pain. *BMC Medicine*. 2017;15(1):35. doi: 10.1186/s12916-016-0775-3
15. Lähdeoja T, Karjalainen T, Jokihaara J, et al. Subacromial decompression surgery for adults with shoulder pain: a systematic review with meta-analysis. *Brit J Sport Med.* 2020;54:665-673. doi: 10.1136/bjsports-2018-100486
16. Hao Q, Devji T, Zeraatkar D, et al. Minimal important differences for improvement in shoulder condition patient-reported outcomes: a systematic review to inform a BMJ Rapid Recommendation. *BMJ Open.* 2019;9:e028777. doi: 10.1136/bmjopen-2018-028777

Chapter 5

A novel framework for incorporating patient values and preferences in making guideline recommendations: guideline panel surveys

Zeng L, Helsingen LM, Bretthauer M, Agoritsas T, Vandvik PO, Mustafa RA, Busse J, Siemieniuk RAC, Lytvyn L, Li SA, Yang M, Yan L, Zhang L, Brignardello-Petersen R, Guyatt GH

Journal of Clinical Epidemiology; 2023; S0895-4356(23)00175-0

Abstract

Objective

Universally acknowledged standards for trustworthy guidelines include the necessity to ground recommendations in patient values and preferences. When information is limited – which is typically the case - guideline panels often find it difficult to explicitly integrate patient values and preferences into their recommendations. Our objective was to develop and evaluate a framework for systematically navigating guideline panels in incorporating patient values and preferences in making recommendations.

Study Design and Setting

In the context of developing a guideline for colorectal cancer screening, we generated an initial framework for creating panel surveys to elicit guideline panelists' views of patient values and preferences and to inform panel discussions on recommendations. With further applications in guidelines of diverse topic areas, we dynamically refined the framework through iterative discussions and consensus.

Results

The finial framework consists of five steps for creating and implementing panel surveys. The surveys can serve three objectives following from the quantitative information regarding patient values and preferences that guideline panels usually require. An accompanying video provides detailed instructions of the survey.

Conclusion

The framework for creating and implementing panel surveys offers explicit guidance for guideline panels considering transparently and systematically incorporating patient values and preferences into guideline recommendations.

Keywords

Patient preferences; Guideline; Survey; Framework; Methods; Threshold

1. Introduction

Standards for the development of trustworthy clinical practice guidelines include the necessity to ground recommendations in patient values and preferences [1,2]. Patient values and preferences are the beliefs, expectations, affinities and priorities for health and life that individuals may apply in considering the potential benefits, harms, burdens and costs of different management options[1]. For guideline developers, patient values and preferences are critical in interpreting research evidence and formulating recommendations [3].

Over the last 25 years, the importance of patient values and preferences in clinical decision-making and in guidelines in particular have been increasingly recognized. In 1999, the JAMA Users' Guides to the medical literature identified the necessity for explicit incorporation of patient values and preferences in clinical practice guidelines [4] and in 2000 another Users' Guide labelled the necessity to consider patient values as a core principle of evidence-based medicine [5]. Both the Cochrane Collaboration and the National Institute for Clinical Excellence (NICE) have placed great emphasis on input from patients and caregivers [6,7]. The GRADE working group has from it's outset emphasized the role of patient values in moving from evidence to decisions [8]. More recently, GRADE has offered an evidence to decision framework that directs guideline developers to examine evidence regarding patient values and preferences and to judge its certainty, as well as considering variability across patients [9,10].

Ideally, cross-sectional surveys among large samples of target patients will be able to inform the relative importance that patients place on different outcomes. Unfortunately, most of the time, such surveys among target population are scarce; even when available, survey results often differ, raising interpretation challenges [11–14]. Consulting with patient partners (i.e., people with lived experience of having the condition or illness, and/or having cared for someone with the condition or illness) or advisory groups, conducting focus group interviews, or reflecting on experience in shared decision making may be helpful, but uncertainty regarding patient values and preferences inevitably remains [15–17]. Hence, guideline panels often need to make inferences of patient values and preferences. Many guidelines fail to make the process of arriving at their inferences regarding values and preferences, or their recommendations, explicit [11,12,14].

In the BMJ Rapid Recommendations - an international clinical practice guideline initiative aiming to produce trustworthy, accessible and timely guidance [18] - we have struggled with appropriately incorporating patient values and preferences. To address the issue our team established a five-step framework for developing and implementing guideline panel surveys to quantitatively ascertain panels' inferences regarding patient values and preferences.

In this article, we describe the development of this framework and illustrate each step within the framework for creating and implementing guideline panel surveys. A paired paper reports the results of a qualitative study evaluating the influence of the surveys in the process of making guideline recommendations [19].

2. Methods

A steering group consisting of experts in guideline methodology and patient values and preferences (GHG, LMH, RAS, POV, TA, LL, MB and LZ) coordinated the development and refinement of the framework for creating and implementing guideline panel surveys.

2.1 Initial development of the framework

In the context of developing recommendations for colorectal cancer screening [20], the steering group constructed a survey for eliciting the guideline panel's view regarding the smallest benefit in colorectal cancer incidence and mortality that, given harms and burdens, the target population would require to undergo screening. Appendix 1 presents a brief introduction to this guideline.

Based on experience from this guideline, the steering group developed an initial framework for using surveys to guide guideline panels in making inferences regarding decision thresholds based on patient values and preferences (e.g., given the harms or burdens of an intervention, what is the smallest benefit patients would require for accepting the intervention). We refer to the overall approach to creating surveys to elicit panel's views regarding patient values and preferences as a "framework" [21].

2.2 Pilot application and refinement of the framework

The steering group applied the framework to another seven guideline panels addressing different topic areas, including the World Health Organization guideline panel addressing therapeutics for COVID-19 [22–28] (Appendix 2). Based on experience with these applications, the steering group, through a process of iterative discussions and consensus, dynamically refined the framework to *i)* extend the objectives of panel surveys, *ii)* finalize and standardize the steps for developing and implementing panel surveys, and *iii)* clarify when guideline panels should consider applying panel surveys (i.e., in which situation panel surveys would be useful).

2.3 Development of an educational video for implementing guideline panel surveys

To facilitate educating guideline panelists, the steering group developed a preliminary version of a video that introduced the key concepts in the panel surveys. Through online user-testing interviews, the steering group collected feedback on the clarity and usefulness of the video. Appendix 3 presents the interview guide for the user-testing interviews.

The steering group anticipated the feedback might vary based on interviewees' prior experience with the panel surveys, and thus using purposeful sampling included guideline panelists with and without experience of taking the panel survey (those who took the panel survey before the development of the video, who did not take the survey and will participate in the survey within the next one to two months, and who did not take the panel survey and do not yet have an explicit plan to apply the survey in the next two months). These interviewees acted as different roles including patient partner, clinical expert, methods co-chair, clinical chair and guideline methodologist from eight different guideline panels.

A professional transcriber transcribed recordings of panel meetings and interviews in English and removed identifying information. Using qualitative description the steering group analysed the transcripts of all interviews in Nvivo™ 12 and refined the video accordingly. Appendix Figure 1 summarizes the development process of the educational video.

The Hamilton Integrated Research Ethics Board (HiREB) approved the evaluation regarding the influence of the guideline panel survey approach on the process of making recommendations (Project Number: 13297) and the user-testing interviews of the educational video (Project Number: 14984).

3. Results

Figure 1 outlines the five-step framework we propose to develop and implement a panel survey directing guideline panelists to make inferences of patient values and preferences. Box 1 illustrates each step using an example.

3.1 Step 1 Judging whether a recommendation is preference-sensitive

The process begins with defining the PICOT (patient, intervention, comparison, outcomes and timeline for measuring the outcomes) of the recommendation. One (usually a steering group of a guideline panel) should consider whether the balance between benefits and harms or burdens is sufficiently close that the recommendation is preference-sensitive. If this is the case, the survey has proved relevant and useful. If the recommendation is not preference-sensitive – in other words, it is clear by the judgement of the steering group or the panel that all or almost all patients would choose or decline the intervention, one need not further consider the survey.

3.2 Step 2 Deciding on survey objective

The objectives of surveys follow from three types of quantitative information regarding patient values and preferences that guideline panels may require:
Objective 1, Establishing the smallest change associated with a single outcome (a benefit or a harm or burden) that patients would perceive as important (minimal important difference, MID);
Objective 2, Given the benefits associated with an intervention, specifying a decision threshold for the maximum key harm or burden that patients would accept for using the intervention; or given the harms or burdens associated with an intervention, specifying a decision threshold for the minimum key benefit that patients would require for using the intervention;

Objective 3, Given best estimates of an intervention's benefits, harms or burdens, making inferences regarding the choices that patients would likely make for or against an intervention.

Box 2 using examples illustrates the three objectives.

3.3 Step 3 Formulating the survey

Achieving Objective 1 and 2 requires specifying a quantified threshold, usually a tough task for panelists. The survey design, acknowledging this challenge, elicits guideline panels' inferences regarding whether patients would perceive a particular magnitude of effect as above or below the underlying MID (Objective 1) or decision threshold (Objective 2). The survey provides a

sequence of magnitudes of effect (the suggested threshold), gradually moving towards an intermediate number (a ping-pong approach going from one extreme to another, gradually narrowing the differences). When a panelist switches his or her response from an effect above the threshold to the effect below the threshold, or vice versa, the panelist effectively identifies a narrow range within which the underlying threshold lies.

For achieving Objective 3, the survey simultaneously presents the effect estimates on benefits and harms or burdens associated with the intervention (usually informed by systematic reviews), and directs panelists to consider whether patients would choose or decline the intervention.

As patient values and preferences differ, the survey asks panelists to infer the distribution of patient values and preferences they would anticipate from a representative group of patients. The standardized options in the survey are as follow: all or almost all (>90%), most (75-90%) or a majority (50-75%) of patients would consider a particular effect as trivial or important (Objective 1), or would choose or decline an intervention (Objective 2 or 3). Box 2 using the examples illustrates the survey designs.

3.4 Step 4 Educating panelists and collecting responses

All guideline panelists including clinicians, content experts, patient partners, guideline methodologists and systematic reviewers can complete the surveys. To prepare panelists for the survey, one may want to consider the video that introduces the key concepts of the survey (https://www.dropbox.com/s/g5pyl7ms5rg7mke/VidePanel%20survey_V10.mp4?dl=0). Extra time to educate patient partners may be advisable.

Through online survey tools, one can collect individual panelists' responses to the survey. To summarize the findings, one can describe the median and the range of panelists' inferences regarding the MID (Objective 1) or the decision threshold (Objective 2) or describe the number of panelists who consider majority of patients would elect for or against the intervention reflecting the panel's inferences regarding the distribution of patients' preferences (Objective 3).

3.5 Step 5 Presenting the findings and eliciting panel discussions

One can present the aggregated findings from the survey in panel meetings to elicit panel's discussions on the interpretation of evidence, the trade-off between benefits and harms, or the direction and strength of recommendations.

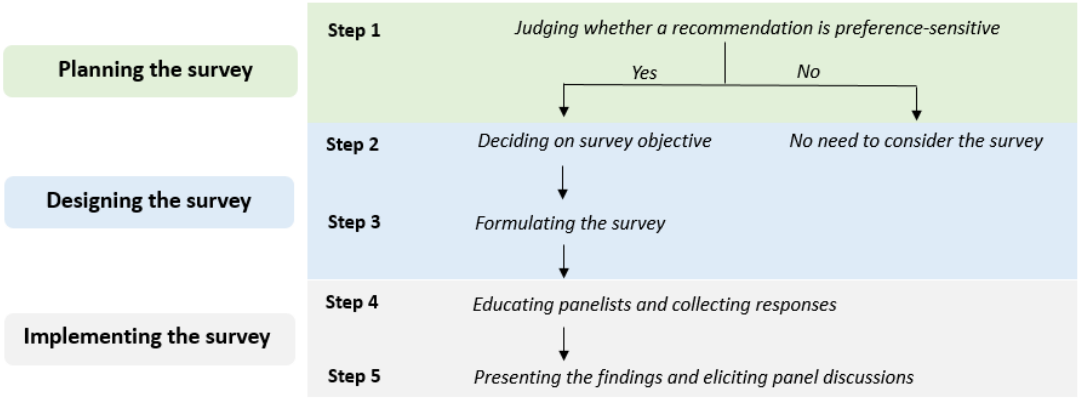


Figure 1 The five-step framework to develop and implement a panel survey directing panelists to make inferences of patient values and preferences

Box 1 An example of applying the five-step framework for developing and implementing a panel survey

Step 1 Judging whether a recommendation is preference-sensitive

Consider a guideline panel making recommendations regarding plasma exchange in addition to usual care (***I***) versus usual care alone(***C***) in patients with ANCA-associated vasculitis (***P***) [24]. The key potential benefit associated with plasma exchange was reduction in end-stage kidney disease (ESKD) (***O***) [24]. The key potential harm was increase in serious infections (***O***). The timeline for measuring both outcomes was one year (***T***).

The steering group of the guideline panel perceived that among patients with ANCA-associated vasculitis, the values and preferences towards plasma exchange varied widely and the survey could thus be useful.

Step 2 Deciding on survey objective

Using data from current trials, the steering group established that the baseline risk of developing ESKD and serious infections varied widely and was strongly associated with the patients' serum creatinine level [29]. Recommendations probably would differ for subgroups of patients with different serum creatinine levels.

To inform the trade-off between the key benefit (reduction in ESKD) and harm or burden (increase in serious infections) associated with plasma exchange, the panel could either establish the minimum benefit that patients would require for accepting plasma exchange (Objective 2), or directly judge the percentage of patients who would elect for or against plasma exchange (Objective 3). As applying Objective 2 would elicit multiple decision thresholds for subgroups, which would require panelists to reflect on more questions, the steering group decided to apply Objective 3.

Step 3 Formulating the survey

The survey for each subgroup of patients presented the baseline risks and corresponding decrease in ESKD and increase in serious infections associated with plasma exchange (informed by a systematic review) [29]. The first scenario was "for patients with serum creatinine level ≤ 200μmol/L, plasma exchange lowers the risk of developing ESKD by 4 in 1000 (from 50 to 46 in 1000), but increases the risk of serious infections by 27 in 1000 (from 100 to 127 in 1000) at 1 year".

The survey asked the panelists "for patients with ANCA-associated vasculitis and with serum creatinine < 200μmol/L, how would patients view the trade-off between the benefit and harm of plasma exchange? "

The options include:

- All or almost all would choose plasma exchange
- Most would choose plasma exchange
- Majority would choose plasma exchange
- Majority would decline plasma exchange
- Most would decline plasma exchange
- All or almost all would decline plasma exchange

In the rest scenarios, the survey presented the benefit and harm associated with plasma exchange in other subgroups (serum creatinine levels at 200-300, 300-400, 400-500, or >500 μmol/L). Following each scenario, the survey asked, given the reduction in ESKD and the increase in serious infections associated with plasma exchange, what proportion of patients would choose or decline plasma exchange. Appendix 4 presents the full survey.

Step 4 Educating panelists and collecting responses

At a panel meeting, the steering group introduced the survey, and had a separate meeting with the patient partners to help them understand the survey. Through an online survey tool, the steering group collected the panelists' responses.

According to these responses, the steering group identified that for patients with serum creatinine level ≤ 300 μmol/L, most panelists perceived that the majority would decline plasma exchange. While for patients with serum creatinine level > 300 μmol/L, most panelists perceived the majority would choose plasma exchange.

Step 5 Presenting the findings and eliciting panel discussions

At the next panel meeting the steering group presented the aggregated findings and launched discussions on the direction and strength of recommendations for subgroups of patients.

Box 2 Examples of three different survey objectives and designs

Objective 1 Establishing an MID threshold

Consider a guideline panel making recommendations for patients with high risk of myocardial infarction regarding a new treatment to reduce that risk [25]. To interpret whether a certain effect of treatment on myocardial infarction is important or not, the panel required information about the smallest reduction in myocardial infarctions that patients would perceive as important (the MID threshold), and thus applied Objective 1.

The survey presented a series of scenarios in which the magnitude of effect on reducing myocardial infarctions varied. The first scenario was "in adults considering the possibility of using the new treatment to reduce the risk of myocardial infarction, the treatment lowers their risk by 1 in 1000 over a period of 5 years". In the following scenarios, the reduction in myocardial infarctions changed to 20, 3, 15, 5, 10, 8 and 12 in 1000 (a ping-pong approach going from one extreme to another, gradually narrowing the differences). Under each scenario, the survey asked panelists to make inferences regarding the proportion of patients who would consider the particular magnitude of effect on myocardial infarction as either important or trivial.

The options include:

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

When a panelist switched the response from "a majority would consider this an important effect" to "a majority would consider this a trivial effect" (or vice versa), the panelist identified a narrow range within which the MID lies. Appendix 5 presents the full surveys for the three examples in Box 2.

Objective 2 Establishing a decision threshold

Consider a guideline panel making recommendations regarding colorectal cancer screening in adults aged 50-79 years [20]. The panel considered reduction in colorectal cancer related mortality as the key benefit and increase in gastrointestinal perforation and major gastrointestinal bleeding as the key harms or burdens.

To tradeoff the key benefit and harms or burdens, the panel required information on the smallest reduction in colorectal cancer related mortality that given harms or burdens people would require to accept screening (the decision threshold), and thus applied objective 2.

Before the panel reviewed evidence on benefit, the survey presented the harms or burdens associated with screening and a series of scenarios in which the absolute reduction in colorectal cancer related mortality varied. The first scenario was "adults screened with colonoscopy have

a 1 in 1000 lower risk of dying from colorectal cancer over a period of 15 years". In the remaining scenarios, the reduction in colorectal cancer related mortality changed to 15, 5 and 10 in 1000 (a ping-pong approach). Following each scenario, the survey asked the panelists to estimate the proportion of adults that would choose or decline screening.

The options include:

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

When a panelist switched the response from "the majority would choose screening" to "the majority would decline screening" (or vice versa), the panelist identified a narrow range within which the decision threshold lies.

Objective 3 Explicitly specifying the percentage of patients who would elect for or against an intervention

Consider a guideline panel making recommendations regarding sodium-glucose transport protein 2 (SGLT 2) inhibitors for patients with type 2 diabetes [23]. The panel considered that the key benefit associated with SGLT 2 inhibitors was reduction in mortality, and the key harms or burdens included increase in genital infection and diabetic ketoacidosis.

Using data from current trials, the panel established that the absolute reduction in mortality associated with SGLT 2 varied widely among patients with different baseline risks [30]. To judge the preferences towards SGLT 2 inhibitors among subgroups of patients, the panel applied Objective 3.

The survey presented the harms or burdens associated with SGLT 2 inhibitors that were constant across subgroups, and then presented the first scenario as "for patients with type 2 diabetes without cardiovascular risk factor (very low risk group), taking SGLT 2 inhibitors has a 5 in 1000 reduction in mortality (from 20 to 15 in 1000) over a period of 5 years". In the remaining scenarios, the reduction in mortality associated with SGLT 2 inhibitors changed to 48, 15, 34 and 5 in 1000 (a ping-pong approach). Following each scenario, the survey asked the panelists to estimate the proportion of patients that would choose or decline SGLT 2 inhibitors. The response reflected panelists' inferences regarding the distribution of choices among subgroups of patients.

4. Discussion

4.1 Main findings and interpretations

When judging the balance between benefits and harms associated with interventions, guideline panels need to interpret the available information and make inferences regarding patient values and preferences that are necessary in moving from evidence to recommendations. We have

developed a novel framework for directing guideline panels to make such inferences, and provide guidance on how those using the framework can develop and implement a panel survey to elicit guideline panelists' view of patient values and preferences.

The panel survey approach allows guideline panels to systematically take the patients' perspective and in doing so make inferences regarding the distribution of patient values and preferences. Incorporating survey findings into the panel discussion clarifies the rationale for panels' decisions regarding the direction and strength of recommendations, thus enhancing the transparency of the process.

The panel survey is not intended to replace primary studies of patient values and preferences (e.g., surveys among patients). Ideally, to optimize panelists' judgments in completing the surveys, practice guidelines will include a review of relevant primary studies (Appendix 5, Example 3 provides an example). Panelists can respond the surveys based on relevant primary studies from such a review, on focus groups commissioned by the guideline panel, on conversations addressing health care decisions with friends or family or, for panelists who are clinicians, on their experience in shared decision-making with patients.

4.2 Strengths and limitations

One prior survey approach, applied in a Chilean COVID-19 living guideline, asked guideline panelists to suggest values of the thresholds for large, moderate, small, or trivial effect [31,32]. The key differences between the survey applied in the Chilean guideline and ours include: we direct guideline panels to think from patients' perspective; surveys are not only applicable for setting thresholds (Objective 1 or 2) but also for directly trading off the benefits versus harms or burdens (Objective 3); recognizing that patient values and preferences often vary, rather than asking panelists to directly specify a threshold or a choice, we ask panelists to infer the distributions of patient values and preferences; finally, we have conducted a qualitative study of the impact of our surveys to inform strengths and limitations [19]. The qualitative evaluation revealed that most panelists found the surveys primed them in considering patient values and preferences and facilitated the incorporation of patient values and preferences in the tradeoffs between benefits and harms or burdens. The variation of patient preferences (provided by responses regarding the distribution of preferences) and uncertainty regarding patient values and preferences (reflected in variation in panelists' responses to the survey questions) helped the panels ponder the strength of recommendations [19]. No other existing approaches provide a formal process for explicitly and systematically interpreting and incorporating patient values and preferences into making recommendations.

One may question guideline panelists' ability to generate insights in patient values and preferences. Indeed, several panelists who participated in our qualitative study raised this issue. Developing recommendations, however, always requires guideline panels to make inferences regarding typical values and preferences – without such inferences, trading off desirable and undesirable consequences of interventions is not possible. Completing the panel survey not only provided best estimates of patient values and preferences, but through variable panel responses, revealed existing uncertainties. Highlighting such uncertainty can inform both the strength of

recommendations (the greater the uncertainty, the more likely a conditional recommendation) and the need for further research regarding values and preferences among target patients.

Studies among patients conducted by investigators associated with two of the guideline panels that participated in our qualitative evaluation (the guideline panel of colorectal cancer screening for adults, and the guideline panel of plasma exchange for ANCA-associated vasculitis) provide some reassurance regarding panel survey results [33,34]. In both cases, although some respondents proved to be uninfluenced to the magnitude of benefits and harms (they chose or declined the intervention across all magnitudes presented), those whose decisions were influenced chose thresholds consistent with panel inferences.

5. Conclusion

When judging the balance between benefits and harms associated with interventions, to formulate recommendations guideline panels must make inferences regarding patient values and preferences. Our proposed framework has proved helpful in facilitating guideline panels' explicit consideration of patient values and preferences and providing an explicit rationale for panels' decisions. We are available for consultation for any guideline panel seeking guidance in creating and implementing panel surveys.

Contributors: All authors made a substantial contribution to the development of the panel survey approach —— they contributed to its development, refinement, and final approval. LNZ, LH, MB, RB-P, GHG developed the initial framework of the panel survey approach. RAVS, TA, POV, LL, RAM, JB tested the panel survey approach in several guidelines. LNZ, S-AL, MTY, LJY, LLZ, RB-P, GHG designed and conducted the qualitative evaluation of the panel survey approach. LNZ and GHG drafted the manuscript. All authors provided feedback and edits on the framework and this paper. All authors approved the final version of the manuscript. LNZ, who led the project, is the guarantor of this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.
Acknowledgments: We thank all the guideline panels that tested and provided feedback on the framework and the panel surveys.

Competing interest: None declared
Funding: No funding was received to develop this panel survey approach. The content reflects the views of it authors.

References

1. Graham R, Mancher M, Wolman DM, et al; Committee on standards for developing trustworthy clinical practice guidelines, board on health care services, Institute of Medicine. Clinical practice guidelines we can trust. 2011.

2. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ*. 2010;182(18):E839-E842. doi:10.1503/cmaj.090449
3. Guidelines International Network PUBLIC Working Group. GIN PUBLIC toolkit: patient and public involvement in guidelines, https://g-i-n.net/new-digital-gin-public-toolkit-patient-and-public-involvement-in-guidelines; 2019 [accessed 18 October 2022].
4. Guyatt GH, Sinclair J, Cook DJ, Glasziou P. Users' guides to the medical literature: XVI. How to use a treatment recommendation. Evidence-Based Medicine Working Group and the Cochrane Applicability Methods Working Group. *JAMA*. 1999;281(19):1836-1843. doi:10.1001/jama.281.19.1836
5. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group. *JAMA*. 2000;284(10):1290-1296. doi:10.1001/jama.284.10.1290
6. Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. *Patient*. 2020;13(2):145-149. doi:10.1007/s40271-019-00408-4
7. Cochrane. Cochrane consumer engagement and involvement framework to 2027, Cochrane consumer engagement and involvement framework to 2027_1.pdf ; 2019 [accessed 13 June 2023].
8. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926. doi:10.1136/bmj.39489.470347.AD
9. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016. doi:10.1136/bmj.i2016
10. Alonso-Coello P, Oxman AD, Moberg J, 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. doi:10.1136/bmj.i2089
11. Blackwood J, Armstrong MJ, Schaefer C, et al. How do guideline developers identify, incorporate and report patient preferences? An international cross-sectional survey. BMC Health Serv Res. 2020;20(1):458. doi:10.1186/s12913-020-05343-x
12. MacLean S, Mulla S, Akl EA, et al. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2 Suppl):e1S-e23S. doi:10.1378/chest.11-2290
13. Lytvyn L, Siemieniuk RA, Dilmitis S, et al. Values and preferences of women living with HIV who are pregnant, postpartum or considering pregnancy on choice of antiretroviral therapy during pregnancy. *BMJ Open*. 2017;7(9):e019023. doi:10.1136/bmjopen-2017-019023
14. Myeloma UK. Measuring patient preferences: an exploratory study to determine how patient preferences data could be used in health technology assessment (HTA), https://www.myeloma.org.uk/wp-content/uploads/2019/07/NICE-Patient-Preferences-Report.pdf s; 2019 [accessed 18 October 2022].

15. Kim C, Berta WB, Gagliardi AR. Exploring approaches to identify, incorporate and report patient preferences in clinical guidelines: Qualitative interviews with guideline developers. *Patient Educ Couns*. 2021;104(4):703-708. doi:10.1016/j.pec.2020.10.001
16. Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11(4):319-329. doi:10.1007/s40258-013-0028-y
17. Food and Drug Administration. Guidance for industry and Food and Drug Administration staff; factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications, https://www.federalregister.gov/documents/2012/03/28/2012-7418/guidance-for-industry-and-food-and-drug-administration-staff-factors-to-consider-when-making; 2012;18828-29 [accessed 18 October 2022].
18. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. *BMJ*. 2016;354:i5191. doi:10.1136/bmj.i5191
19. Zeng L, Li SA, Yang M, et al. Qualitative study of guideline panelists: Innovative surveys provided valuable insights regarding patient values and preferences. *J Clin Epidemiol*. 2023;S0895-4356(23)00186-5. doi:10.1016/j.jclinepi.2023.07.014
20. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;367:l5515. doi:10.1136/bmj.l5515
21. Helsingen LM, Zeng L, Siemieniuk RA, et al. Establishing thresholds for important benefits considering the harms of screening interventions. *BMJ Open*. 2020;10(12):e037854. doi:10.1136/bmjopen-2020-037854
22. Word Health Organization. Therapeutics and COVID-19: living guideline, https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-therapeutics-2022.2; 2022 [accessed 21 April 2022].
23. Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ*. 2021;373:n1091. doi:10.1136/bmj.n1091
24. Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ*. 2022;376:e064597. Published 2022 Feb 25. doi:10.1136/bmj-2021-064597
25. Hao Q, Aertgeerts B, Guyatt G, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations. *BMJ*. 2022;377:e069066. doi:10.1136/bmj-2021-069066
26. Ye Z, Busse JW, Hill MD, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: A clinical practice guideline. *J Evid Based Med*. 2022;15(3):263-271. doi:10.1111/jebm.12493
27. Conservative, pharmacological and surgical interventions for patients with temporomandibular disorder: a clinical practice guideline. 2023 [Forthcoming].
28. Interventional procedures for axial and radicular chronic spine pain: a clinical practice guideline. 2023 [Forthcoming].
29. Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: an updated systematic review and meta-analysis. *BMJ*. 2022;376:e064604. doi:10.1136/bmj-2021-064604

30. Palmer SC, Tendal B, Mustafa RA, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ*. 2021;372:m4573. doi:10.1136/bmj.m4573
31. Neumann I, Quiñelen E, Nahuelhual P, et al. Using Explicit Thresholds were valuable for judging Benefits and Harms in partially contextualized GRADE Guidelines. *J Clin Epidemiol*. 2022;147:69-75. doi:10.1016/j.jclinepi.2022.03.017
32. Morgano GP, Mbuagbaw L, Santesso N, et al. Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomised methodological study (GRADE-THRESHOLD). *BMJ Open*. 2022;12(3):e053246. doi:10.1136/bmjopen-2021-053246
33. Usher-Smith JA, Mills KM, Riedinger C, et al. The impact of information about different absolute benefits and harms on intention to participate in colorectal cancer screening: A think-aloud study and online randomised experiment. *PLoS One*. 2021;16(2):e0246991. doi:10.1371/journal.pone.0246991
34. Collister D, Farrar M, Farrar L, et al. Plasma Exchange for ANCA-Associated Vasculitis: An International Survey of Patient Preferences. *Kidney Med*. 2022;5(3):100595. doi:10.1016/j.xkme.2022.100595

Appendix 1 Brief introduction of BMJ *Rapid recommendations* for colorectal cancer screening in adults

At the time the recommendations were published, recent 15-year updates of sigmoidoscopy screening trials provide new evidence on the effectiveness of colorectal cancer screening [1]. Prompted by the new evidence, the guideline panel asked: "Does colorectal cancer screening make an important difference to health outcomes in individuals initiating screening at age 50 to 79? And which screening option is best?" [1].

The guideline adhered to standards developed for *BMJ Rapid Recommendations*. These standards do not compromise on the rigor of either the systematic reviews or other aspects of the development of trustworthy guidelines, but endeavor to complete each step in guideline development as quickly as possible.

The guideline panel consisted of 22 panelists, including members of the public with experience with colorectal cancer screening, clinicians, colorectal cancer screening experts and guideline methodologists.

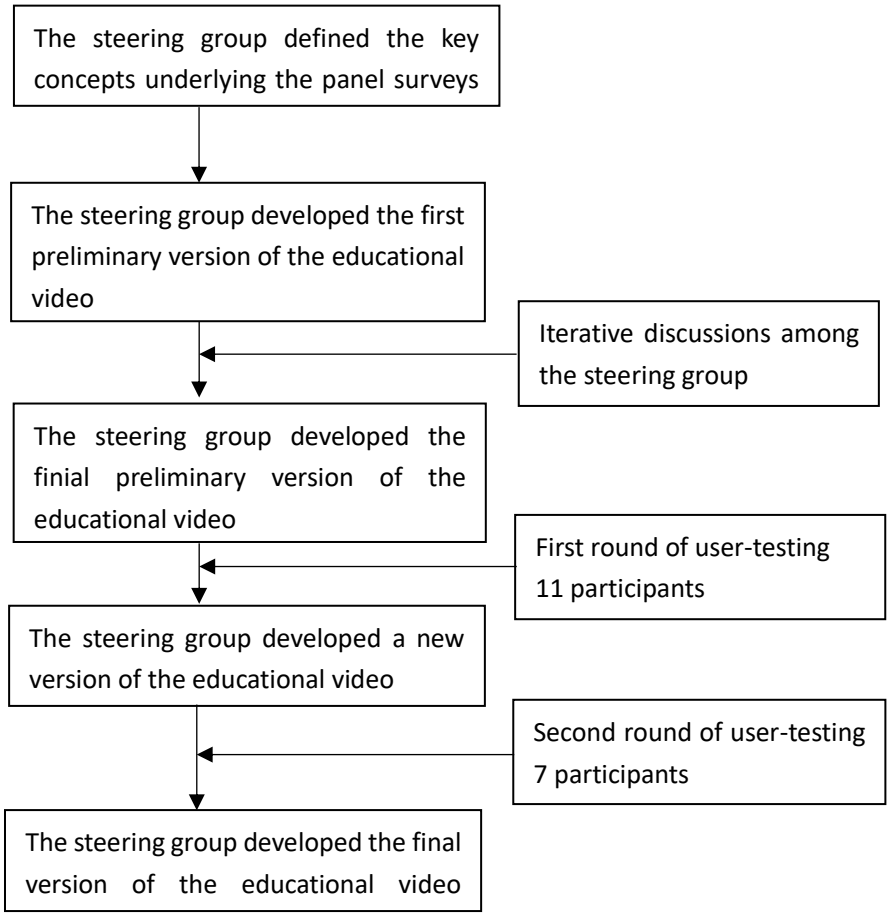
For understanding patient values and preferences, the panel first conducted a systematic review of quantitative estimates of the thresholds of benefits and harms required for individuals to choose to undergo screening. However, with limited data available the systematic review failed to provide useful information. The guideline panel then performed two panel surveys to elicit panelists' inferences regarding the smallest reduction in mortality associated with screening (i.e., the key benefit) that target population would require for accepting screening over no screening or for choosing one screening test over others.

Appendix 2 Topic areas of the seven guidelines that inform the development and refinement of the framework

Guideline	Target population	Intervention and comparison	Key outcome of interest	Objective applied
Therapeutics and COVID-19: living guideline [2]	Patients with COVID-19	Drug therapies in addition to usual care versus usual care, or one drug therapy versus another	Mortality, mechanical ventilation, admission to hospital, time to symptom resolution, adverse effects leading to drug discontinuation *etc*.	Objective 1
SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline [3]	Patients with type 2 diabetes	SGLT-2 inhibitors in addition to usual care versus usual care, GLP-1 receptor agonists in addition to usual care versus usual care, or SGLT-2 inhibitors versus GLP-1 receptor agonists	All-cause mortality, cardiovascular death, myocardial infarction, end stage kidney disease, serious hyperglycaemia *etc*.	Objective 3
Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline [4]	Patient with ANCA-associated vasculitis	Plasma exchange in addition to usual care versus usual care	Mortality, end stage kidney disease, serious infections *etc*.	Objective 2,3
PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations [5]	Patients with elevated low-density lipoprotein (LDL) cholesterol using high dose statins or intolerant to statins	Adding a second lipid-lowering drug versus not adding, adding ezetimibe versus PCKS 9 inhibitor first	All-cause mortality, cardiovascular mortality, non-fatal myocardial infarction, non-fatal stroke *etc*.	Objective 1
Endovascular thrombectomy and intravenous	Patients with acute ischemic stroke due to large vessel	Endovascular thrombectomy and intravenous	Minimal disability measured by	Objective 1

alteplase in patients with acute ischemic stroke due to large vessel occlusion: a clinical practice guideline [6]	occlusion	alteplase Intravenous alteplase in combination with endovascular thrombectomy (EVT) versus EVT alone	modified Rankin Score 0–2, mortality, symptomatic intracranial hemorrhage *etc.*	
Conservative, pharmacological and surgical interventions for patients with temporomandibular disorder: a clinical practice guideline [7]	Adults presenting with chronic pain associated with a temporomandibular disorder	Pharmacological and surgical interventions versus placebo or usual care, or one intervention versus another	Pain intensity, physical function *etc*.	Objective 2
Interventional procedures for axial and radicular chronic spine pain: a clinical practice guideline [8]	Adults with axial and radicular spine pain	Interventional procedures versus placebo or usual care, or one interventional procedure versus another	Pain intensity, physical function *etc.*	Objective 2

Appendix Figure 1 The development process of the educational video



Appendix 3 Interview guide for the user-testing of the educational video

Instruction

- During the interview, I will walk you through the video, and will stop very frequently so that you can share your impressions of bite-sized sections of the video.

[General impression of the video]

1. Before we start the interview, do you have any question about the panel survey exercise, or what we are about to do going through the interview?
2. Any chance you watched the video or part of it before the interview?
   Before we go through the video together, any general comment, you would like to share, on the video?

[Content and format of the video]

3. How would you describe the **clarity and usefulness** of this segment of the video? (The interviewer asked this question for each bite sized piece that captures the presentation of a single key concept of the survey approach).
4. We are just switching from Part 1 to Part 2, is the transition clear? (The interview will ask this question at the switch point)
   Probes:
   The video has four parts:
   Part I: the role of values and preferences in making recommendations (no right or wrong; concern values and preferences of target population; need to consider the distribution of values and preferences);
   Part II: the approaches for understanding patient values and preferences (directly asking the patients; think from the target populations' perspective; think about the distribution of values and preferences);
   Part III: examples of the panel survey (clarifying three objectives of the panel survey approach).
   Part IV: Other key issues of the panel survey approach that need to clarify (thinking from target population's perspective is difficult but necessary; panel will consider factors other than benefits and harms or burdens later in the panel discussions; survey can help understand panel member's opinions and help panel members discuss recommendations).

[Other comments of the video]

5. Beyond what you have commented, is there anything missing from the video? If so what?
6. Beyond what you have commented, is there anything can be omitted? If so what?
7. Overall, how would you think the coherence of the video?
8. Overall, how would you think the acceptability of the video to the guideline panel members?
   Probe: the length, the presentation (more flashing animation?)
9. Is there any other experience about the video you would like to share with us?

Thank you so much for your time today. Closing as appropriate.

Appendix 4 A panel survey of values and preferences towards plasma exchange in patients with ANCA-associated vasculitis

Introduction

Purpose of this survey: We would like to know the panel's perspective of the distribution of choices individuals would make after full shared decision-making regarding whether or not to use plasma exchange. We will use your responses to inform our discussion of the tipping point, with regard to baseline serum creatinine levels, at which the majority would switch from declining to accepting plasma exchange.

Content of this survey: We will present you the key benefit (absolute risk reduction in end-stage kidney disease) and harm or burden (absolute risk increase in serious infections) of plasma exchange in patients with ANCA-associated vasculitis and with baseline serum creatinine at ≤200μmol/Lμmol/L, 200-300μmol/Lμmol/L, 300-400μmol/Lμmol/L, 400-500μmol/L, >500μmol/L. We will then ask you for your perspective about what proportion of patients would choose or decline plasma exchange under each scenario. Each question will vary the risk reduction of end-stage kidney disease and the risk increase of serious infections. Please read these carefully.

6. For patients with ANCA-associated vasculitis and with **serum creatinine ≤ 200μmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **4 in 1000 lower risk** of end-stage kidney disease at 1 year (from 50 to 46 in 1000)
Harms: **27 in 1000 increased risk** of serious infections at 1 year (from 100 to 127 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

7. For patients with ANCA-associated vasculitis and with **serum creatinine > 500μmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **168 in 1000 lower risk** of end-stage kidney disease at 1 year (from 400 to 232 in 1000)
Harms: **135 in 1000 increased risk** of serious infections at 1 year (from 500 to 635 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

8. For patients with ANCA-associated vasculitis and with **serum creatinine 200-300μmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **31 in 1000 lower risk** of end-stage kidney disease at 1 year (from 75 to 44 in 1000)

Harms: **49 in 1000 increased risk** of serious infections at 1 year (from 180 to 229 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

9. For patients with ANCA-associated vasculitis and with **serum creatinine <u>400-500µmol/L</u>**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **104 in 1000 lower risk** of end-stage kidney disease at 1 year (from 275 to 171 in 1000)
Harms: **97 in 1000 increased risk** of serious infections at 1 year (from 360 to 457 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

10. For patients with ANCA-associated vasculitis and with **serum creatinine <u>300-400µmol/L</u>**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **57 in 1000 lower risk** of end-stage kidney disease at 1 year (from 150 to 93 in 1000)
Harms: **73 in 1000 increased risk** of serious infections at 1 year (from 270 to 343 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

Appendix 5 Full panel surveys of the three examples in Box 2

Example 1 A survey of values and preferences among patients with risk of myocardial infarctions (Objective 1 Establishing an MID threshold)

Introduction

Based on the last panel survey, we identified the reduction in myocardial infarction as a key desirable outcome. Now we need your help in considering what impact on the outcome that patients would consider important. This will help us in rating precision when deciding on certainty of evidence and will ultimately help in deciding on recommendations for or against the drugs under consideration, and the strength of those recommendations.

We need you to answer a series of questions regarding what patients would consider a trivial or important effect. At this point, the question is abstract in that it isn't tied to the benefits, harms, or burdens of interventions. These judgments are challenging. If you are a clinician, please reflect the question based on your experience in shared decision-making with your patients. If you are attending the panel in the role of patient, please answer the question on the basis of conversations with friends, family, and acquaintances around making health care decisions.

In the following questions, when we say "all or almost all", we mean 90% or more; when we say most, we mean 75% to 90%; and when we say the majority, we mean 50% to 74% (also applicable to other examples in Appendix 3).

In each case, the proportion who think an effect is important will be 100% minus the proportion who think an effect is trivial. For instance, if you choose the option "the majority would consider this a trivial effect" it means that you think that 50% to 74% would think the effect trivial and 26% to 49% would think it is important.

1. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **1 in 1000** (i.e. a decrease in myocardial infarction of 1 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

2. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **20 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that

will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

3. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **3 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

4. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **15 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

5. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **5 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect

- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

6. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **10 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

7. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **8 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

8. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **12 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

Example 2 A survey of values and preferences towards colorectal cancer screening in adults aged 50-79 years (Objective 2 Establishing a decision threshold)

Introduction

The purpose of the following exercise is to make you think through the questions and provide a common understanding of the magnitude of effects. This survey will give us an idea of where people stand that would help us with the upcoming discussions. We will present you the integrated answers from the panel before we start discussion in our next panel meeting. Please do not consider your answers here as final. You should all be open to change your minds as the discussion progresses.

We may make different recommendations for colorectal cancer screening (in particular colonoscopy) for people with different risks of colorectal cancer and colorectal cancer death. We might recommend against screening for people with a low absolute risk of being diagnosed with or dying from colorectal cancer, and recommend for screening in people at higher absolute risk. Therefore, we need to establish the threshold where our recommendations shift. Please do the following exercise with focus on the population – what would most people do if they were fully informed of the potential burden related to screening.

Harms and burdens associated with colonoscopy

Procedure-related mortality: Fewer than 1 per 1000 procedures
Gastrointestinal perforations: Approximately 1 per 1000 procedures
Major gastrointestinal bleedings: Approximately 3 per 1000 procedures

1. Adults aged 50-79 years screened with colonoscopy have a **1 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

2. Adults aged 50-79 years screened with colonoscopy have a **15 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

3. Adults aged 50-79 years screened with colonoscopy have a **5 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

4. Adults aged 50-79 years screened with colonoscopy have a **10 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

Example 3 A survey of values and preferences towards SGLT2 inhibitor in patients with type 2 diabetes (Objective 3 Explicitly specifying the percentage of patients who would elect for or against an intervention)

Introduction

We would like to know the panel's views about the degree of benefit and harm that is important for patients, to inform our discussion about recommendations.

We will present the results of a focus group study about values and preferences of people living with type 2 diabetes, and a values and preferences review. We will also present to you the benefits and harms of SGLT2 inhibitor for each of the 5 risk groups, and ask you for your view about what proportion of patients would choose the drug or standard care.

We will use the results of the focus group, values and preferences review, and this survey to inform our guideline recommendations discussion. Please consider the following descriptions of benefit and harm outcomes when answering the questions in the survey. Consider a 5 year timeframe for all outcomes.

Summary of focus group study

Methods

- 7 participants (6 male, 1 female) living with type 2 diabetes, based in Canada
- First exercise - looking at individual benefit outcomes of SGLT2 inhibitors (reduction in risk of mortality, end-stage renal disease, myocardial infarction, heart failure requiring hospitalization, and stroke):

For each drug, when presented with all its harms, asked how much of a reduction in risk for each outcome would make it worthwhile to choose the drug

- Second exercise - looking at all outcomes of SGLT2 inhibitors:

For each drug, when presented with all its harms and benefits, for each of the 5 risk groups, asked if they would accept the drug over standard care

Results from individual outcomes exercise

- For any outcomes, about a third of participants were willing to accept very small benefit (chose less than 5 in 1000 reduction in risk), a third were not willing to accept the largest possible benefit (would not choose even more than 30 in 1000 reduction in risk), and a third varied in their threshold (chose between 5 and 30 in 1000 reduction in risk)

- Risk thresholds contradicted additional discussion about medication choice:

If blood sugar is managed, most participants not willing to take additional medications

Participants valued short term outcomes of harm more than long term outcomes of Benefit

- Reduction in risk of end stage renal disease was weighed similarly to mortality

- Reduction in risk of myocardial infarction, stroke, and heart failure requiring hospitalization were weighed about the same, but less than end stage renal disease and mortality

- Injection medications are less desirable than oral medications*

*Similarly, a systematic review on values and preferences found that oral drugs are preferable to injection drugs.

Harms of SGLT2 inhibitors

Please consider the following harms, which are constant among all risk groups:

Risk of diabetic ketoacidosis = from 2 in 1000 to 4 in 1000 (2 in 1000 increase)

Risk of genital infection = from 73 in 1000 to 212 in 1000 (139 in 1000 increase)

1. For patients with type 2 diabetes without cardiovascular risk factor (**very low risk**), how would patients view such effects?

All-cause mortality = 20 in 1000 to 15 in 1000 (**5 in 1000 reduction**)

Patient focus group decision: 2 yes (28%), 5 no (72%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

2. For patients with type 2 diabetes with 3 or more cardiovascular risk factors (**low risk**), how would patients view such effects?

All-cause mortality = 70 in 1000 to 55 in 1000 (**15 in 1000 reduction**)

Patient focus group decision: 3 yes (43%), 4 no (57%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

3. For patients with type 2 diabetes with established cardiovascular disease (**moderate risk**), how would patients view such effects?

All-cause mortality = 120 in 1000 to 95 in 1000 (**25 in 1000 reduction**)

Patient focus group decision: 4 yes (57%), 3 no (43%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

4. For patients with type 2 diabetes with established chronic kidney disease (**high risk**), how would patients view such effects?

All-cause mortality = 170 in 1000 to 136 in 1000 (**34 in 1000 reduction**)

Patient focus group decision: 4 yes (57%), 3 no (43%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

5. For patients with type 2 diabetes with established cardiovascular disease and chronic kidney disease (**very high risk**), how would patients view such effects?

All-cause mortality = 265 in 1000 to 217 in 1000 (**48 in 1000 reduction**)

Patient focus group decision: 5 yes (72%), 2 no (28%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

References

1. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;367:l5515. doi:10.1136/bmj.l5515
2. Word Health Organization. Therapeutics and COVID-19: living guideline, https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-therapeutics-2022.2; 2022 [accessed 21 April 2022].
3. Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ*. 2021;373:n1091. doi:10.1136/bmj.n1091
4. Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ*. 2022;376:e064597. doi:10.1136/bmj-2021-064597
5. Hao Q, Aertgeerts B, Guyatt G, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations. *BMJ*. 2022;377:e069066. doi:10.1136/bmj-2021-069066
6. Ye Z, Busse JW, Hill MD, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: A clinical practice guideline. *J Evid Based Med*. 2022;15(3):263-271. doi:10.1111/jebm.12493
7. Conservative, pharmacological and surgical interventions for patients with temporomandibular disorder: a clinical practice guideline. 2023 [Forthcoming].
8. Interventional procedures for axial and radicular chronic spine pain: a clinical practice guideline. 2023 [Forthcoming].

Chapter 6

Qualitative study of guideline panelists: Innovative surveys provided valuable insights regarding patient values and preferences

Zeng L, Li SA, Yang M, Yan L, Helsingen LM, Bretthauer M, Agoritsas T, Vandvik PO, Mustafa RA, Busse J, Siemieniuk RAC, Lytvyn L, Zhang L, Brignardello-Petersen R, Guyatt GH

Journal of Clinical Epidemiology; 2023; S0895-4356(23)00186-5

Abstract

Objective: To explore guideline panelists' understanding of panel surveys for eliciting panels' inferences regarding patient values and preference, and the influence of the surveys on making recommendations.

Study design and setting: We performed sampling and data collection from all four guideline panels that had conducted the surveys through October 2020. We collected the records of all panel meetings, and interviewed some panelists in different roles. We applied inductive thematic analysis for analyzing and interpreting data.

Results: We enrolled four guideline panels with 99 panelists in total, and interviewed 25 of them. Most panelists found the survey was easy to follow and facilitated the incorporation of patient values and preferences in the tradeoffs between benefits and harms or burdens. The variation of patient preferences and uncertainty regarding patient values and preferences reflected in the surveys helped the panels ponder the strength of recommendations. In doing so, the survey results enhanced a rationale for panels' decision on the recommendations.

Conclusion: The panel surveys have proved to help guideline panels explicitly consider and incorporate patient values and preferences in making recommendations. Guideline panels would benefit from widespread use of the panel surveys, particularly when primary evidence regarding patient values and preferences is scarce.

Keywords

Patient preferences; Guideline; Recommendation; Survey; Interpretive description; Qualitative study

1. Introduction

Clinical practice guidelines are statements that include recommendations intend to optimize patient care. To make a recommendation, guideline panels should define clinical questions, select the relevant outcome variables, retrieve and synthesize all the relevant evidence, rate confidence in the effect estimates and, relying on a systematic approach but ultimately also on consensus, move from evidence to recommendations [1]. Guideline panels often provide detailed information about the process they followed in developing a guideline, however, how they incorporate patient values and preferences in trading off the benefits and harms associated with interventions often remains untransparent [2].

Guideline panels' neglect of explicit consideration of patient values and preferences is understandable: panelists are sometimes unfamiliar with the process, and relevant research is often scarce or completely unavailable, and if available challenging to interpret [3–6]. Thus, guideline panels often need to make inferences of patient values and preferences. These challenges may not be problematic if benefits far outweigh harms and burdens or the reverse. But when the balance is closer, careful judgment of the relative importance of outcomes becomes critical [7, 8].

A research team developed a framework for creating and implementing guideline panel surveys to facilitate guideline panelists making inferences regarding patient values and preferences. The team using the framework implemented surveys for guideline panels in different topic areas, including the World Health Organization (WHO) guideline panel addressing therapeutics for COVID-19 [9–15]. A paired paper reports in detail the development of this framework and illustrate each step within the framework for creating and implementing guideline panel surveys [16]. Box 1 introduces the key characteristics of the panel surveys.

Box 1 Panel surveys for facilitating guideline panelists making inferences regarding patient values and preferences

1. Objectives

The objectives of the panel surveys follow from three types of quantitative information regarding patient values and preferences that guideline panels may require: establishing thresholds for the minimally important difference (Objective 1, the smallest change associated with a single outcome that patients would perceive as important); establishing decision thresholds (Objective 2, patients' choice of accepting or declining an intervention would reverse when the effect associated the intervention falls on one side or another of the threshold); or explicitly judging whether benefits of interventions outweigh harms or burdens (Objective 3).

2. Participants

The surveys are not intended for patients but rather for guideline panelists. All panelists including clinicians, content experts, methodologists and patient partners (i.e., people with lived experience of having the condition or illness, and/or having cared for someone with the condition or illness) or public panelists (i.e., the general public) can complete the surveys.

3. How the surveys work

The surveys ask guideline panelists to think from the patients' perspective and considering the distribution of patient values and preferences.

4. The relationship between the surveys and other sources of information on patient values and preferences

Rather than replacing other methods for understanding patient values and preferences, the surveys aim to provide a structured and transparent approach for guideline panels to make inferences regarding patient values and preferences. Prior to taking the surveys, panelists should review any studies available that address patients' values and preferences (e.g., surveys of patients, focus group commissioned by the guideline panel).

2. Methods

2.1 Study design

Using interpretive description (i.e., an inductive qualitative method that allows for the exploration of individual and shared experiences across contexts) [17], a research team with expertise in qualitative research, guideline methodology, and patient values and preferences conducted the qualitative evaluation. The Hamilton Integrated Research Ethics Board approved this project (Project Number: 13297). The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist guided the approach to reporting [18].

2.2 Sampling and data collection

We conducted sampling and data collection in two phases. In the first phase, we performed sampling and data collection from involved members of all four guideline panels that had applied the framework for conducting surveys through October 2020. These four guideline panels are the first four panels that applied the panel survey approach [9, 11–13].

All these four guidelines are *BMJ Rapid Recommendations* [19] developed under a collaboration between the *MAGIC Evidence Ecosystem Foundation* (a non-profit research and innovation program) and the *BMJ*. The guideline steering group, with representatives from *MAGIC* and the *BMJ*, chose experts with research experience on the basis of the steering group's familiarity with the relevant literature. They chose practicing clinicians on the basis of personal contacts or a snowballing approach starting with personal contacts. They chose methodologists from personal contacts and patient partners from patient organizations. The patient partners, like other panelists on the panels for *BMJ Rapid Recommendations*, participated in the finalization of the questions, considering patient values and preferences, the tradeoff between benefits and harms, and the finalization of recommendations. The guideline panels followed principles of trustworthy guideline development as implemented in *BMJ Rapid Recommendations* [19]. Panel chairs introduced the issue of values and preferences and its importance in trading off desirable and undesirable consequences of interventions, and in particular how uncertainties and variability in values and preferences might influence strength of recommendations. Panels then considered, and in some cases undertook, systematic reviews of relevant literature regarding values and preferences.

Panels then discussed their impressions of typical patient values and preferences, and impressions of the associated uncertainty and variability. When, in these discussions, it became evident that a structured approach to eliciting panel views on these matters was desirable, they undertook the panel surveys.

Box 2 briefly introduces the context of each guideline and the methods the guideline panels used for understanding patient values and preferences. Appendix 1 presents the panel surveys applied by each panel.

Box 2 Context of the four sample guidelines and the methods the guideline panels used for understanding patient values and preferences

Guideline 1 Colorectal cancer screening for adults

This guideline made recommendations regarding colorectal cancer screening for adults aged 50-79 years [9].

To understand patient values and preferences, the panel first conducted a systematic review of relevant studies. The limited data available failed to provide useful information. To elicit panelists' inferences regarding the smallest reduction in mortality associated with screening that patients would demand to undergo screening or to choose one test over another, the guideline panel then conducted the panel survey.

Guideline 2 Sodium-glucose co-transporters-2 (SGLT-2) inhibitors and glucagon-like peptide 1 (GLP-1) receptor agonists for patients with type 2 diabetes

This guideline made recommendations regarding SGLT 2 inhibitors and GLP-1 receptor agonists for patients with type 2 diabetes [11].

The panel first conducted a systematic review on values and preferences that provided limited information. The panel then performed a focus group study of seven participants living with type 2 diabetes addressing the participants' values and preferences. Informed by the focus group, the panel performed the panel survey regarding given the benefits (e.g., reduction in all-cause mortality) and harms (e.g., increase in genital infection or diabetic ketoacidosis) associated with the interventions, what proportion of patients would choose or decline the interventions or would choose one intervention over another.

Guideline 3 Plasma exchange for antineutrophil cytoplasmic antibodies (ANCA)-associated vasculitis

This guideline made recommendations regarding plasma exchange for patients with ANCA-associated vasculitis [12].

The panel identified subgroups of patients for whom the key benefit (i.e., reduction in end-stage kidney disease) and the key harm associated with plasma exchange (i.e., increase in serious infections) varied. For each subgroup, given the magnitude of reduction in end-stage kidney disease and the magnitude of increase in serious infections, the survey asked guideline

panelists to estimate the proportion of patients that would choose or decline plasma exchange.

Guideline 4 Proprotein convertase subtilisin/kexin-9 (PCSK-9) inhibitors for patients with dyslipidemia

This guideline panel made recommendations regarding PCSK 9 inhibitors for patients with dyslipidemia and/or at cardiovascular risk [13].

To set a threshold of the minimal important reduction in myocardial infarction, and thus to inform precision ratings for this key outcome, the guideline panel conducted a panel survey. The survey framed the question as the smallest change associated with myocardial infarction that patients would perceive as important.

In the second phase, aiming to simultaneously capture a wide range of perspectives from panelists with different roles and the opinions from guideline panelists who expressed strong concerns regarding the panel surveys in panel meetings, we used a combination of maximum variation sampling and extreme sampling strategies to sample individual guideline panelists from each guideline panel. Through a virtual conference platform (Zoom) [20], an interviewer with qualitative methods training (LZ) conducted one-to-one interviews with guideline panelists' in different roles from the four guideline panels to understand panelists' experiences of applying the surveys. Appendix 2 presents the interview guide. Because, in both phases, we achieved thematic saturation - defined as findings in later stages of the data interpretation yielding confirmation of earlier findings but nothing new [21] - with the first four guideline panels, we enrolled no further panels.

We obtained written consent for using anonymized recordings of panel meetings and interviews from all panelists. A professional transcriber transcribed recordings of panel meetings and interviews in English and removed identifying information. Nvivo™ 12 was used to organize and store text, and to support coding and data interpretation [22].

2.3 Data analysis and interpretation

Data analysis and interpretation occurred concurrently with data collection. We applied inductive thematic analysis, using data to generate codes for analyzing and interpreting data [23–25]. One coder (LZ) conducted the initial rounds of analysis, through analysis of six transcripts of panel meetings and six transcripts of interviews, developed initial codebooks for panel meetings and for individual interviews. A core team (LZ, S-AL, GHG, RB-P) met periodically to share perspectives and to develop consensus on the codes and underlying themes. Another two coders (LY, MY) with graduate preparation in a health profession and specialized training in qualitative research, using the refined codebooks, independently coded 25% of the transcripts and cross compared with the first coder. The core team discussed new findings beyond the codebook and resolved discrepancies in coding.

After coding all relevant panel meetings and interviews, we organized together data segments that reflected similar thematic patterns. We looked for commonalities and differences between the thematic patterns, and collapsed associated thematic patterns into themes (Figure 1). Frequently returning to the primary passages ensured that the emergent themes were grounded in the data [26]. As relationships between the themes became apparent, we conceptualized the findings by extracting key themes representing the guideline panelists' understanding of the surveys, their experience of the influence of the surveys on the process of making recommendations. To be transparent with our description, when reporting the findings from individual interviews, if more than 90% of the participants commented on a thematic pattern we reported as "almost all of the participants"; if 50%-90%, "majority or most"; 10%-50%, "minority or some"; less than 10% "few or none". The selection is these thresholds is arbitrary but could help increase the transparency of the reporting. We shared a synthesis of findings from panel meetings and interviews and asked four randomly selected panelists who participated in the interviews to reflect on the credibility of the interpretation (i.e. member checking).

2.4 Rigor and reflexivity

The inter-professional composition of clinicians, content experts, guideline methodologists, systematic review leaders and patient partners in the interviews helps capture different perspectives. The triangulation of data sources (i.e. panel meetings and individual interviews) and member checking contribute to verify the findings.

To identify forces that might skew the research in particular directions, at the pre-research stage, we, as the researchers of this qualitative evaluation study and developers of the panel survey approach, examined our motivations, assumptions and interest in this study [27]. We met to discuss what we thought we were likely to find in this study. This reflexivity allowed us to become aware of our presuppositions regarding the impact and experience of the panel surveys and helped us not to impose them on our interviewees.

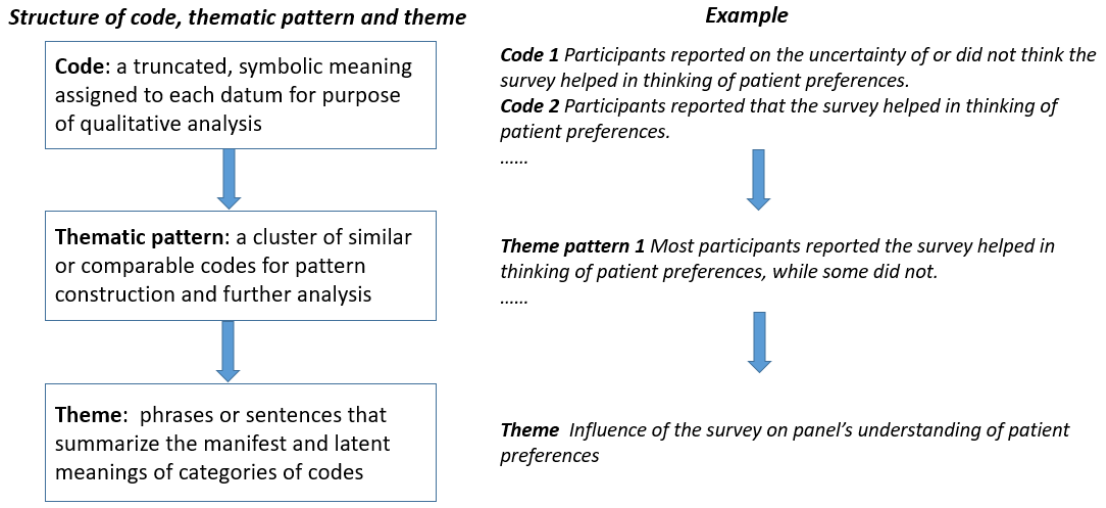


Figure 1 An example of the structure of code, thematic pattern and theme

3. Results

We enrolled four guideline panels with 99 panelists in total, and invited 30 of them to the interviews. Among the 30 invited we finally interviewed 25 participants including four acted as the role of chairs or methods co-chairs, 11 as clinical experts or content experts, five as guideline methodologists or systematic reviewers, and five as patient partners. The rest five panelists did not respond our invitation or refused due to lack of time. Appendix Table 1 reported the characteristics of guideline panels and panelists.

Six themes emerged from the analysis of the panel meetings and individual interviews: panelists' understanding of the panel surveys; the influence of surveys on panels' understanding of patient values and preferences; on their tradeoff between benefits and harms; on their decisions on recommendations; as well as the panels' comments on the challenges and limitations of the surveys.

3.1 Theme 1 Panels' understanding of the panel surveys

The majority of the participants found the survey was framed in a "straightforward fashion" and "easy to follow". Most understood that the survey required them to reflect from the target population's perspective. The majority clearly understood that the survey asked to consider the distribution of patient values and preferences while the minority expressed "it is not easy to get that point" *(Clinical Expert and Methodologist 11, Interview)*.

3.2 Theme 2 Influence of the surveys on guideline panels' understanding of patient values and preferences

Most participants in the interviews reflected that the survey primed them in considering patient values and preferences, especially in a quantitative way rather than a vague or unspecified manner. During panel meetings, the survey prompted discussions regarding patient values and preferences, with chairs reviewing the survey results and leading the panels to consider the central tendency, variation and the panel's uncertainty regarding patient values and preferences (Appendix Box 1).

3.3 Theme 3 Influence of the surveys on guideline panels' trade-off between benefits and harms or burdens of interventions

Most participants felt the survey facilitated the panel in considering trade-offs between benefits and harms or burdens associated with the interventions. The survey focused panelists on considering the same issues in making trade-offs: "the survey helped us to ... [be] in the same mindset of envisioning seeing exactly the kind of trade-off that we are talking about" *(Clinical Chair 2, Interview)*. The survey reminded the panel to incorporate patients' perspective into the trade-off: "all of the panelists were very sensitive to the findings of the survey... the entire panel did take a moment to reflect on what patients would want before coming up with a final trade-off" *(Clinical Expert 34, Interview)*.

3.4 Theme 4 Influence of the surveys on guideline panels' discussions regarding the direction and strength of recommendations

The majority of the participants felt that the surveys informed the panel discussions on recommendations. By providing fellow panelists' thoughts and the average opinions of the panel on patient values and preferences, the surveys prompted panelists to consider their judgement on the recommendations: "The averaging out of the opinions [from the survey] led to something that was acceptable to most people in the group" *(Clinical Expert 22, Interview)*. By revealing variation or uncertainty of the panel's view on patient values and preferences, the survey informed the panel discussions on the strength of recommendations: "Half of the panel was for strong recommendation, the other half with it for a weak [also called conditional] recommendation. And that was a moment that we often did come back to this panel survey saying, we did see that there was a large variability… in light of that fact we should try to strive for a weak recommendation" *(Clinical Expert 13, Interview)*. The influence of the surveys on recommendations was also evident from the panel discussions (Appendix Box 2).

Panelists felt the survey was particularly valuable when subgroups existed, in which the balance between benefits and harms or burdens was close: "This [the survey] was a good way to structure the discussion around subgroups... If you thought about it [the subgroups] early enough, the survey is just crystallizes this…but if you haven't thought about it early enough, it will force you to commit to specific subgroups, which is important" *(Clinical Expert 27, Interview)*. The influence of the surveys on illustrating the impact of subgroups was also evident from the panel discussions. (Appendix Box 2, Objective 3, *Clinical Expert 27, Panel meeting #3 Alternative*).

A minority of participants, who disagreed with the survey results, found the approach unhelpful in making recommendations. Most such concerns arose from members of the Colorectal Cancer Screening Guideline panel: "I didn't give much value to the threshold. Because honestly in clinics it's not that relevant" *(Clinical Expert 32, Interview)*.

3.5 Theme 5 Influence of the surveys on the efficiency and transparency of making recommendations

Having panels' views before panel meetings, chairs reflected that the surveys helped them to structure the panel discussions: "as the chair, [I need] to be very prepared [with regards] to where should the discussion focus on" *(Clinical Chair 1, Interview)*. Through focused discussions on the main trade-offs, the majority of the participants reported that the survey improved the efficiency of panel discussions: "it really tightened the discussion… it was very good for efficiency, making the meetings run quite smoothly" *(Methodologist 5, Interview)*. The survey also helped the panels achieve consensus on recommendations: "the panelists when they come across this survey they would figure out that these are the results … So, they can perhaps discuss amongst themselves and come to reach a consensus that these are the values that patients would like to have and who perhaps should go for a recommendation according to their values" *(Clinical Expert and Methodologist 11, Interview).*

Some expressed that providing panel's view on a pre-specified threshold, or how patients would trade-off the benefits and harms, the survey improved the transparency of the process of making recommendations. The surveys provided a forum in which panelists could express opinions

independently, and avoided “the loud person who might try to drag the conversation in one direction” *(Methodologist 5, Interview)*.

3.6 Theme 6 Challenges and limitations of the surveys

When we asked the participants about whether it was easy or difficult to take the perspective of patients as described in the survey, the majority reported “it is not easy to become the representative of the patient” *(Clinical Chair 2, Interview)*. They expressed uncertainty as to what extent they could represent patients, and identified some barriers.

Small sample size and lack of representativeness of the panel was a major concern for some participants: “I would argue that having a larger sample size might give you more accurate results” *(Clinical Expert 10, Interview)*. A few participants expressed that patients in different settings might have differing values and preferences and panelists with different expertise would have diverse views on patient values and preferences.

Some panelists expressed concerns that factors not presented in the surveys might influence patients’ decisions: “it [the factors in the survey] might not be the only things that would influence a patients’ decision… the discussion that needs to come up what other things are being presented and what other things are the patient's facing when making that decision” *(Patient Partner 4, Interview)*.

4. Discussion

4.1 Main findings and interpretations

In this article, we evaluated guideline panelists’ understanding of the panel surveys and their experiences regarding the influence of the panel surveys on the process of making guideline recommendations. The qualitative evaluation revealed that most panelists found the surveys were easy to follow and primed them in considering patient values and preferences and facilitated the incorporation of patient values and preferences in the tradeoffs between benefits and harms or burdens. The variation of patient preferences (provided by responses regarding the distribution of preferences) and uncertainty regarding patient values and preferences (reflected in variation in panelists’ responses to the survey questions) helped the guideline panels ponder the strength of recommendations. In doing so, the survey results enhanced a rationale for guideline panels’ decision on the recommendations.

Not all panelists were enthusiastic regarding the surveys, and this was particularly the case for those who disagreed with the majority view. One might argue that not all guideline panelists have the ability to generate insights in patient values and preferences [28,29]. Nevertheless, developing recommendations requires guideline panels to make inferences regarding typical values and preferences – without such inferences, trading off desirable and undesirable consequences of interventions is impossible. Completing the survey revealed existing uncertainties regarding values and preferences, and for some this uncertainty undermined the usefulness of the survey. Highlighting uncertainties, however, can inform both the strength of recommendations (the greater the uncertainty, the more likely a conditional recommendation) and the need for further research regarding values and preferences among target patients.

Guideline panelists pointed out that the survey focused only on benefits, harms or burdens, and ignored other factors in the evidence to decision framework. Indeed, resource requirements, acceptability, feasibility and equity concerns may have important influence on guideline panels' recommendations [30]. Importantly, the chosen guidelines used to develop surveys took an individual patient perspective, not a healthcare systems perspective. By highlighting the panels' views regarding patient values and preferences, the surveys inform the always crucial discussion of benefits, harms and burdens; this does not preclude, and may facilitate clarity in the discussion of other important factors in decision-making.

4.2 Strengths and limitations

In our study, the inter-professional composition of the guideline panelists including clinicians, methodologists (including systematic review leaders) and patient partners facilitated a breadth of perspectives in evaluating the potential usefulness of the panel surveys. The triangulation of data sources (i.e., responses to panel surveys, recordings of guideline panel meetings and individual interviews) also contributed to the overall rigor of this study [24]. We provided a detailed audit trail to defend decisions made during the research process, and included the participants' narrative within the findings to demonstrate the quality of the research findings (i.e. prolonged engagement) [31].

A potential limitation of our study is that the guideline panels involved in this study were recruited through key informants, and the panels used similar methods in making guideline recommendations (e.g., using the GRADE for rating the certainty of evidence, and considering the direction and strength of recommendations). To address this, we drew on guideline panelists with varying levels of experience, and with different attitudes towards the panel surveys to capture potential diverse experiences from guideline panelists.

5. Conclusion

The panel surveys have proved to prime guideline panels to make inferences regarding patient values and preferences, a fundamental responsibility of guideline panels, and to help guideline panels explicitly consider and incorporate patient values and preferences in making recommendations. The results of this qualitative study suggested that many guideline panels would benefit from widespread use of the panel surveys, particularly when primary evidence regarding patient values and preferences was scarce.

Contributors: All authors made a substantial contribution to the evaluation of the influence of panel surveys. LNZ, S-AL, MTY, LJY, LLZ, RB-P, GHG designed and conducted the qualitative evaluation of the panel survey approach. LNZ and GHG drafted the manuscript. All authors provided feedback and edits on the framework and this paper. All authors approved the final version of the manuscript. LNZ, who led the project, is the guarantor of this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgments: We thank all the guideline panelists who gave us permission to analyze the panel meeting records, and also thank the participants who shared their experience of applying the panel survey approach in the interviews.

Competing interest: None declared
Funding: This project is supported by the Einstein Foundation Berlin as part of the Einstein Foundation Award for Promoting Quality in Research. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, the Einstein Foundation or the award jury.

References

1. Neumann I, Akl EA, Vandvik PO, et al. How to use a patient management recommendation: clinical practice guidelines and decision analyses. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice, 3rd ed. New York, NY: McGraw-Hill Education; 2015.
2. Van der Wees PJ, Hendriks EJ, Custers JW, Burgers JS, Dekker J, de Bie RA. Comparison of international guideline programs to evaluate and update the Dutch program for clinical guideline development in physical therapy. *BMC Health Serv Res*. 2007;7:191. doi:10.1186/1472-6963-7-191
3. Blackwood J, Armstrong MJ, Schaefer C, et al. How do guideline developers identify, incorporate and report patient preferences? An international cross-sectional survey. *BMC Health Serv Res*. 2020;20(1):458. doi:10.1186/s12913-020-05343-x
4. MacLean S, Mulla S, Akl EA, et al. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2 Suppl):e1S-e23S. doi:10.1378/chest.11-2290
5. Lytvyn L, Siemieniuk RA, Dilmitis S, et al. Values and preferences of women living with HIV who are pregnant, postpartum or considering pregnancy on choice of antiretroviral therapy during pregnancy. *BMJ Open*. 2017;7(9):e019023. doi:10.1136/bmjopen-2017-019023
6. Myeloma UK. Measuring patient preferences: an exploratory study to determine how patient preferences data could be used in health technology assessment (HTA), https://www.myeloma.org.uk/wp-content/uploads/2019/07/NICE-Patient-Preferences-Report.pdf s; 2019 [accessed 18 October 2022].
7. Brett Hauber A, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11(4):319-329. doi:10.1007/s40258-013-0028-y
8. Food and Drug Administration. Guidance for industry and Food and Drug Administration staff; factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications,

https://www.federalregister.gov/documents/2012/03/28/2012-7418/guidance-for-industry-and-food-and-drug-administration-staff-factors-to-consider-when-making; 2012;18828-29 [accessed 18 October 2022].
9. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;367:l5515. doi:10.1136/bmj.l5515
10. Word Health Organization. Therapeutics and COVID-19: living guideline, https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-therapeutics-2022.2; 2022 [accessed 21 April 2022].
11. Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ*. 2021;373:n1091. doi:10.1136/bmj.n1091
12. Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ*. 2022;376:e064597. . doi:10.1136/bmj-2021-064597
13. Hao Q, Aertgeerts B, Guyatt G, et al. PCSK9 inhibitors and ezetimibe for the reduction of cardiovascular events: a clinical practice guideline with risk-stratified recommendations. *BMJ.* 2022;377:e069066. doi:10.1136/bmj-2021-069066
14. Ye Z, Busse JW, Hill MD, et al. Endovascular thrombectomy and intravenous alteplase in patients with acute ischemic stroke due to large vessel occlusion: a clinical practice guideline. *J Evid Based Med*. 2022;15(3):263-271. doi:10.1111/jebm.12493
15. Conservative, pharmacological and surgical interventions for patients with temporomandibular disorder: a clinical practice guideline. 2023 [Forthcoming].
16. Zeng L, Helsingen LM, Bretthauer M, et al. A novel framework for incorporating patient values and preferences in making guideline recommendations: guideline panel surveys. *J Clin Epidemiol*. 2023;S0895-4356(23)00175-0. doi:10.1016/j.jclinepi.2023.07.003
17. Thorne, S. Interpretive Description: Qualitative Research for Applied Practice (2nd ed.). Routledge. 2016. doi:10.4324/9781315545196
18. Booth A, Hannes K, Harden A, et al. COREQ (Consolidated Criteria for Reporting Qualitative Studies). Guidelines for Reporting Health Research: A User's Manual2014. p. 214-26.
19. Siemieniuk RA, Agoritsas T, Macdonald H, Guyatt GH, Brandt L, Vandvik PO. Introduction to BMJ Rapid Recommendations. *BMJ.* 2016;354:i5191. doi:10.1136/bmj.i5191
20. Zoom Video Communications Inc. Security guide. San Jose, CA: Zoom Video Communications Inc. https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf; 2016) [accessed 14 February 2022].
21. Leavy P. The Oxford handbook of qualitative research. 2nd ed. Oxford University Press, USA; 2020.
22. QSR International. NVivo qualitative data analysis Software Version 11.1. 1, https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home; 2015 [accessed 14 February 2022].
23. MacQueen, K. M., McLellan, E., Kay, K., & Milstein, B. Codebook Development for Team-Based Qualitative Analysis. CAM Journal. 1998; 10(2): 31–36. doi:10.1177/1525822X980100020301
24. Krefting L. Rigor in qualitative research: the assessment of trustworthiness. *Am J Occup Ther.* 1991;45:214-22. doi:10.5014/ajot.45.3.214.

25. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol:* 2006;3:77-101. doi:10.1191/1478088706qp063oa.
26. Giacomini MK, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: XXIII. Qualitative research in health care. B. What are the results and how do they help me care for my patients? *JAMA.* 2000;284(4):478–482. doi:10.1001/jama.284.4.478.
27. Polkinghorne DE. Phenomenological research methods. In: R. Valle & S. Halling, editors. Existential-phenomenological perspectives in psychology: Exploring the breadth of human experience, Springer Science & Business Media; 1989, p. 41-60.
28. Ubel PA, Angott AM, Zikmund-Fisher BJ. Physicians recommend different treatments for patients than they would choose for themselves. *Arch Intern Med.* 2011;171:630-34. doi:10.1001/archinternmed.2011.91.
29. Mulley AG, Trimble C, Elwyn G. Stop the silent misdiagnosis: patients' preferences matter. *BMJ*. 2012;345:e6572. doi:10.1136/bmj.e6572.
30. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016. doi:10.1136/bmj.i2016.
31. Finlay L. "Outing" the researcher: the provenance, process, and practice of reflexivity. *Qual Health Res.* 2002;12:531-45. doi:10.1177/104973202129120052.

Appendix 1 Panel surveys applied by the four sample guideline panels

Guideline 1

Panel survey regarding values and preferences towards colorectal cancer screening in adults aged 50-79 years (Objective 2 Establishing a decision threshold)

Introduction

The purpose of the following exercise is to make you think through the questions and provide a common understanding of the magnitude of effects. This survey will give us an idea of where people stand that would help us with the upcoming discussions. We will present you the integrated answers from the panel before we start discussion in our next panel meeting. Please do not consider your answers here as final. You should all be open to change your minds as the discussion progresses.

We may make different recommendations for colorectal cancer screening (in particular colonoscopy) for people with different risks of colorectal cancer and colorectal cancer death. We might recommend against screening for people with a low absolute risk of being diagnosed with or dying from colorectal cancer, and recommend for screening in people at higher absolute risk. Therefore, we need to establish the threshold where our recommendations shift. Please do the following exercise with focus on the population – what would most people do if they were fully informed of the potential burden related to screening.

Harms and burdens associated with colonoscopy

Procedure-related mortality: Fewer than 1 per 1000 procedures
Gastrointestinal perforations: Approximately 1 per 1000 procedures
Major gastrointestinal bleedings: Approximately 3 per 1000 procedures

5. Adults aged 50-79 years screened with colonoscopy have a **1 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

6. Adults aged 50-79 years screened with colonoscopy have a **15 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

7. Adults aged 50-79 years screened with colonoscopy have a **5 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

8. Adults aged 50-79 years screened with colonoscopy have a **10 in 1000** (0.1%) lower risk of dying from colorectal cancer at 15 years. How would they view such benefits?

- All or almost all would choose screening
- Most would choose screening
- A majority would choose screening
- A majority would decline screening
- Most would decline screening
- All or almost all would decline screening

Guideline 2

Panel survey regarding values and preferences towards SGLT2 inhibitor in patients with type 2 diabetes (Objective 3 Explicitly specifying the percentage of patients who would elect for or against an intervention)

Introduction

We would like to know the panel's views about the degree of benefit and harm that is important for patients, to inform our discussion about recommendations.

We will present the results of a focus group study about values and preferences of people living with type 2 diabetes, and a values and preferences review. We will also present to you the benefits and harms of SGLT2 inhibitor for each of the 5 risk groups, and ask you for your view about what proportion of patients would choose the drug or standard care.

We will use the results of the focus group, values and preferences review, and this survey to inform our guideline recommendations discussion. Please consider the following descriptions of benefit and harm outcomes when answering the questions in the survey. Consider a 5 year timeframe for all outcomes.

Summary of focus group study

Methods

- 7 participants (6 male, 1 female) living with type 2 diabetes, based in Canada

- First exercise - looking at individual benefit outcomes of SGLT2 inhibitors (reduction in risk of mortality, end-stage renal disease, myocardial infarction, heart failure requiring hospitalization, and stroke):

For each drug, when presented with all its harms, asked how much of a reduction
In risk for each outcome would make it worthwhile to choose the drug

- Second exercise - looking at all outcomes of SGLT2 inhibitors:

For each drug, when presented with all its harms and benefits, for each of the 5 risk groups, asked if they would accept the drug over standard care
Results from individual outcomes exercise

- For any outcomes, about a third of participants were willing to accept very small benefit (chose less than 5 in 1000 reduction in risk), a third were not willing to accept the largest possible benefit (would not choose even more than 30 in 1000 reduction in risk), and a third varied in their threshold (chose between 5 and 30 in 1000 reduction in risk)

- Risk thresholds contradicted additional discussion about medication choice:

If blood sugar is managed, most participants not willing to take additional medications
Participants valued short term outcomes of harm more than long term outcomes of Benefit

- Reduction in risk of end stage renal disease was weighed similarly to mortality

- Reduction in risk of myocardial infarction, stroke, and heart failure requiring hospitalization were weighed about the same, but less than end stage renal disease and mortality

- Injection medications are less desirable than oral medications*

*Similarly, a systematic review on values and preferences found that oral drugs are preferable to injection drugs.

Harms of SGLT2 inhibitors

Please consider the following harms, which are constant among all risk groups:
Risk of diabetic ketoacidosis = from 2 in 1000 to 4 in 1000 (2 in 1000 increase)
Risk of genital infection = from 73 in 1000 to 212 in 1000 (139 in 1000 increase)

6. For patients with type 2 diabetes without cardiovascular risk factor (**very low risk**), how would patients view such effects?

All-cause mortality = 20 in 1000 to 15 in 1000 (**5 in 1000 reduction**)

Patient focus group decision: 2 yes (28%), 5 no (72%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors

- All or almost all would decline SGLT2 inhibitors

7. For patients with type 2 diabetes with 3 or more cardiovascular risk factors (**low risk**), how would patients view such effects?

All-cause mortality = 70 in 1000 to 55 in 1000 (**15 in 1000 reduction**)

Patient focus group decision: 3 yes (43%), 4 no (57%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

8. For patients with type 2 diabetes with established cardiovascular disease (**moderate risk**), how would patients view such effects?

All-cause mortality = 120 in 1000 to 95 in 1000 (**25 in 1000 reduction**)

Patient focus group decision: 4 yes (57%), 3 no (43%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

9. For patients with type 2 diabetes with established chronic kidney disease (**high risk**), how would patients view such effects?

All-cause mortality = 170 in 1000 to 136 in 1000 (**34 in 1000 reduction**)

Patient focus group decision: 4 yes (57%), 3 no (43%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

10. For patients with type 2 diabetes with established cardiovascular disease and chronic kidney disease (**very high risk**), how would patients view such effects?

All-cause mortality = 265 in 1000 to 217 in 1000 (**48 in 1000 reduction**)

Patient focus group decision: 5 yes (72%), 2 no (28%)

- All or almost all would choose SGLT2 inhibitors
- Most would choose SGLT2 inhibitors
- Majority would choose SGLT2 inhibitors
- Majority would decline SGLT2 inhibitors
- Most would decline SGLT2 inhibitors
- All or almost all would decline SGLT2 inhibitors

Guideline 3

Panel survey regarding values and preferences towards plasma exchange in patients with ANCA-associated vasculitis (Objective 3 Explicitly specifying the percentage of patients who would elect for or against an intervention)

Introduction

Purpose of this survey: We would like to know the panel's perspective of the distribution of choices individuals would make after full shared decision-making regarding whether or not to use plasma exchange. We will use your responses to inform our discussion of the tipping point, with regard to baseline serum creatinine levels, at which the majority would switch from declining to accepting plasma exchange.

Content of this survey: We will present you the key benefit (absolute risk reduction in end-stage kidney disease) and harm or burden (absolute risk increase in serious infections) of plasma exchange in patients with ANCA-associated vasculitis and with baseline serum creatinine at ≤200µmol/Lµmol/L, 200-300µmol/Lµmol/L, 300-400µmol/Lµmol/L, 400-500µmol/L, >500µmol/L. We will then ask you for your perspective about what proportion of patients would choose or decline plasma exchange under each scenario. Each question will vary the risk reduction of end-stage kidney disease and the risk increase of serious infections. Please read these carefully.

11. For patients with ANCA-associated vasculitis and with **serum creatinine ≤ 200µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **4 in 1000 lower risk** of end-stage kidney disease at 1 year (from 50 to 46 in 1000)
Harms: **27 in 1000 increased risk** of serious infections at 1 year (from 100 to 127 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

12. For patients with ANCA-associated vasculitis and with **serum creatinine > 500µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **168 in 1000 lower risk** of end-stage kidney disease at 1 year (from 400 to 232 in 1000)
Harms: **135 in 1000 increased risk** of serious infections at 1 year (from 500 to 635 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

13. For patients with ANCA-associated vasculitis and with **serum creatinine 200-300µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **31 in 1000 lower risk** of end-stage kidney disease at 1 year (from 75 to 44 in 1000)
Harms: **49 in 1000 increased risk** of serious infections at 1 year (from 180 to 229 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

14. For patients with ANCA-associated vasculitis and with **serum creatinine 400-500µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **104 in 1000 lower risk** of end-stage kidney disease at 1 year (from 275 to 171 in 1000)
Harms: **97 in 1000 increased risk** of serious infections at 1 year (from 360 to 457 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

15. For patients with ANCA-associated vasculitis and with **serum creatinine 300-400µmol/L**, how would patients view the trade-off between the benefits and harms of plasma exchange?

Benefits: **57 in 1000 lower risk** of end-stage kidney disease at 1 year (from 150 to 93 in 1000)
Harms: **73 in 1000 increased risk** of serious infections at 1 year (from 270 to 343 in 1000)

- All or almost all (over 90%) would **choose** plasma exchange
- Most (75- 90%) would **choose** plasma exchange
- Majority (51-74%) would **choose** plasma exchange
- Majority (51-74%) would **decline** plasma exchange
- Most (75-90%) would **decline** plasma exchange
- All or almost all (over 90%) would **decline** plasma exchange

Guideline 4

Panel survey regarding values and preferences among patients with risk of myocardial infarctions (Objective 1 Establishing an MID threshold)

Introduction

Based on the last panel survey, we identified the reduction in myocardial infarction as a key desirable outcome. Now we need your help in considering what impact on the outcome that patients would consider important. This will help us in rating precision when deciding on certainty of evidence and will ultimately help in deciding on recommendations for or against the drugs under consideration, and the strength of those recommendations.

We need you to answer a series of questions regarding what patients would consider a trivial or important effect. At this point, the question is abstract in that it isn't tied to the benefits, harms, or burdens of interventions. These judgments are challenging. If you are a clinician, please reflect the question based on your experience in shared decision-making with your patients. If you are attending the panel in the role of patient, please answer the question on the basis of conversations with friends, family, and acquaintances around making health care decisions.

In the following questions, when we say "all or almost all", we mean 90% or more; when we say most, we mean 75% to 90%; and when we say the majority, we mean 50% to 74% (also applicable to other examples in Appendix 3).

In each case, the proportion who think an effect is important will be 100% minus the proportion who think an effect is trivial. For instance, if you choose the option "the majority would consider this a trivial effect" it means that you think that 50% to 74% would think the effect trivial and 26% to 49% would think it is important.

9. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **1 in 1000** (i.e. a decrease in myocardial infarction of 1 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

10. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **20 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

11. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **3 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

12. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **15 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

13. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **5 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

14. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **10 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

15. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **8 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

16. In adults considering the possibility of using medication to reduce their risk of myocardial infarction, an intervention lowers their risk by **12 in 1000** (i.e. a decrease in myocardial infarction of 15 in 1,000 patients) over a period of 5 years. Please choose an option that will reflect the proportion of patients who would think this reduction in risk is an (a) important/trivial effect?

- All or almost all would consider this an important effect
- Most would consider this an important effect
- A majority would consider this an important effect
- A majority would consider this a trivial effect
- Most would consider this a trivial effect
- All or almost all would consider this a trivial effect

Appendix 2 Interview guide for semi-structured interviews with guideline panelists

The purpose of this interview is to explore the influence of the formal panel survey approach on the process of making recommendations. We are going to record this interview. Recording allows us to look back on your thoughts and feelings to better understand your experience, and we can make the interview data more helpful. In future reports we may include direct quotes to help describe the study findings; however, each interview will be de-identified, meaning that anything said will not be linked back to the person who made the statement. We don't anticipate that there are any risks associated with your participation, but you have the right to stop the interview at any time or skip any question that you do not want to answer.

[Presenting the interviewee with the survey questionnaire]

The interviewer will present the survey questionnaire to remind you about the survey questions and options.

[Understanding of the survey approach]

1. What are your general thoughts about the survey exercise(s)?
2. When reflecting on the survey, did you realize the survey asked to think from the patients' perspective? Do you think it is easy or difficult to do so?
3. When reflecting on the survey, did you realize the survey asked to consider the distribution the patient preferences? Do you think it is easy or difficult to do so?

[Influence of the survey]

4. How would you describe the influence of the survey exercise(s), if any, on getting YOU think about people's values and preferences?
5. How would you describe the influence of the survey exercise(s), if any, on getting YOU to think of the TYPICAL people's values and preferences?
6. How would you describe the influence of the survey exercise(s), if any, on getting YOU to think of the DISTRIBUTION of people's values and preferences?

[For Objective 1 and 2]

7. How would you describe if any, the influence of this pre-specified threshold on YOU when you traded off the benefits and harms of *[the intervention]*?

[For Objective 3]

8. Thinking of the typical people's values and preferences with different levels of pairs of benefits and harms of *[the intervention] (i.e. reduction/increase in+- [the key benefit or harm]),* how would you describe if any, the influence of the survey findings on YOU when you traded off the benefits and harms of *[the intervention]*?

[Hereafter for all Objectives]

9. When you consider recommendations for *[the intervention]* during the panel meetings, did the survey results in any way influence YOUR judgement of the direction or strength of recommendations?

10. How do you describe if any, the influence of the survey on knowing the other panelists' thoughts about patient preferences?
11. How do you describe, if any, the influence of the survey on the efficiency of the process of making recommendations?
12. Overall, how useful do you think the survey exercise is? In which situations, you think the panel survey would be helpful?
13. What is your thoughts about how the panel should use the survey findings?
14. What downsides, disadvantages or limitation of using the survey if any?
15. What suggestions do you have for improving the survey exercise?
16. Is there any other experience about the survey exercise you would like to share with us?

Thank you so much for your time today. Closing as appropriate.

Appendix Table 1 Characteristics of guideline panels and panelists

Characteristic	Number (%)
Type of panel surveys applied by the guideline panels*	PCSK 9 guideline panel: Objective 1 Colorectal(CR) Cancer Screening guideline panel: Objective 2 SGLT-2 guideline panel: Objective 3 Vasculitis guideline panel: Objective 2,3
Number of panel meetings relevant to panel surveys/Total Number of panel meetings (%)	13/20 (6%)
Number of panelists participated in the interviews/Total Number of panelists from the four guideline panels**	25/99 (25%)
Chairs, methods co-chairs	4/13 (31%)
Clinical experts or content experts	11/43 (26%)
Guideline methodologists, systematic review leaders	5/26 (19%)
Patient partners	5/17 (29%)
Number of female/Total Number of panelists (%)	40/99 (40%)
Number of female in interviews/Total Number of panelists in interviews (%)	11/25 (44%)

Abbreviation: SGLT-2= Sodium-glucose transport protein 2; GLP-1= Glucagon-like peptide-1; PCSK= Proprotein convertase subtilisin/kexin; ANCA= Antineutrophil cytoplasmic antibody

* Objective 1 Establishing the smallest change associated with a single outcome (a benefit or a harm or burden) that patients would perceive as important (minimal important difference, MID);

Objective 2 Given the benefits associated with an intervention, specifying a decision threshold for the maximum key harm or burden that patients would accept for using the intervention; or given the harms or burdens associated with an intervention, specifying a decision threshold for the minimum key benefit that patients would require for using the intervention;

Objective 3 Given best estimates of an intervention's benefits, harms or burdens, making inferences regarding the choices that patients would likely make for or against an intervention.

** Panelists include chairs, co-chairs, chair mentors and panel members.

Appendix Box 1 Influence of the panel surveys on guideline panels' understanding of patient values and preferences

Thematic pattern 1 The survey primed panels in thinking of patient values and preferences

"the survey exercise is very useful not only in terms of providing context but thought process… it really allowed you to put yourself in the patient's perspective in that moment that they are trying to make that decision"*(Patient Partner 4, Interview)*.

"So first it forced me myself to put numbers on what before was more or less an intuition…the survey was one of the steps that helped me to get my ideas clearer than they were before" *(Clinical Chair 2, Interview)*.

Thematic pattern 2 The survey prompted panel discussions regarding the central tendency of patient values and preferences

"The result of that [survey] question indicates that the average value of the benefit of plasma exchange in terms of decreasing end stage kidney disease should be 4.5%. So, a 4.5% benefit would sort of justify 6% increase in terms of harms and in terms of risk of serious infection" *(Clinical Chair 2, Panel meeting #1)*.

"I am going to… show you the results for all of the risk groups. Now when we go to the low risk groups with some risk factors the majority flips. However, there are still people who think maybe not, we shouldn't be giving it [SGLT2 inhibitors]. When we go to the moderate risk groups, so now it is people with cardiovascular disease, same thing, the majority says give it [SGLT2 inhibitors], but the number is higher." *(Clinical Chair 1, Panel meeting #5)*.

Thematic pattern 3 The survey prompted panel discussions regarding the variation or uncertainty of patient values and preferences

"We have very different notions about what the majority of people, were they fully informed, would choose. Some of us think that even two or three in a thousand would make people inclined to use the intervention. And some of us think that it would require twenty in a thousand for the majority to choose the interventions…And the bottom line perhaps not surprisingly our inferences differ widely in terms of what we think typical values and preferences are." *(Methods Co-Chair Mentor 1, Panel meeting #1)*.

"Now, 95% think of us think that almost or majority would choose screening. 5% however still think that majority would decline" *(Clinical Chair 4, Panel meeting #3)*.

Appendix Box 2 Influence of the panel surveys on panels' discussions regarding the direction and strength of recommendations

Thematic pattern 1 Referring to the survey results, the panels discussed the direction of recommendations

Objective 1

Referring to the MID threshold, the chair led the panel to discuss the direction of recommendations: "Let's see the PCSK9 added to statins versus statins alone, the benefits here is a little bit a little bit bigger but it's also a trivial, because it's less than our preset minimum important difference. For nonfatal MI, it's 5 fewer per thousand, and for nonfatal stroke is 6 fewer per thousand, and the certainty of evidence is moderate. So considering this, do we want to add a second drug for this population?" (Methods Co-Chair 2, Panel meeting #3).

Referring to the MID threshold, the panelists expressed their opinions on the direction of recommendations: "I feel like we remind ourselves about the [MID] thresholds we want to use to guide our decisions. And the confidence intervals here, they clearly overlap no effect. I think it would be very hard to justify a weak recommendation for Ezetimibe… the best estimate effect is six fewer under the [MID] threshold right?" *(Clinical Expert and Methodologist 4, Panel meeting #2)*.

Objective 2

Referring to the decision threshold, the chair led the panel to discuss the direction of recommendations: "Colonoscopy gets to the threshold of [reduction in colorectal cancer incidence of] 10 in a thousand [at the baseline risk of] a little bit less than 3%. Sigmoidoscopy gets there a little bit above 3%...Since that is where it crosses our threshold of 10 in a thousand reduction, above 3% we would be recommending in favour of colonoscopy and sigmoidoscopy. And below three and a thousand we would not." *(Clinical Chair 4, Panel meeting #6)*.

Referring to the decision threshold, the panelists expressed opinions on the direction of recommendations: "I think for me the concept is that tests … which meet the threshold for both outcomes may have a more favourable of set of attributes than tests which meet only one of the two." *(Clinical Expert 28, Panel meeting #4)*.

Objective 3

Referring to panel's judgement of typical patients' trade-off, the chair led the panel to discuss the direction of recommendations: "based on the panel survey for the patient with creatinine of less than 200, 80% said voted against plasma exchange in this group, and 20% said still in favour of plasma exchange in this group. So, probably based on the panel survey indicating a weak against of plasma exchange in this group" *(Methods Co-Chair 4, Panel meeting #3)*.

Referring to panel's judgement of typical patients' trade-off, the panelists expressed opinions on the direction of recommendations for subgroups: "I when I look at here [the survey results], so clearly showing different creatinine level as an indicator truly for different risk groups…I do agree that it appears that [creatinine level of] 300 is like threshold where the benefits and harm differ… the 3 [300] to 400 group is hard group … But we are clearly putting more weight on the end-

stage kidney disease compared to the serious infections. Um, so I agree with that recommendation, weak or suggest against [plasma exchange] for [patients with creatinine level] below 300. And suggest for, for more than 300" *(Clinical Expert 27, Panel meeting #3 Alternative)*.

Thematic pattern 2 Referring to the survey results, the panels discussed the strength of recommendations

"From the variability of the [survey] results shown here strongly suggests that all of our recommendations should be weak because we have very different views about the values and the preferences in the population."*(Methods Co-Chair 2, Panel meeting #1)*.

"I think the other major point that I wanted to make is that since there is variability in our panel in the survey that speaks to weak recommendations rather than strong recommendations." *(Clinical Expert and Methodologist 6, Panel meeting #3)*.

Chapter 7

General discussion and future perspectives

This thesis starts with the development of a clinical practice guideline aiming to transfer the latest evidence regarding plasma exchange and reduced dose of corticosteroids in patients with ANCA-associated vasculitis into clinical practice (Chapter 2) and discussed two methodological issues raised from evidence synthesis and evidence to decision process of this guideline (Chapter 3-6). This concluding chapter reviews the key findings of these studies and explores challenges as well as opportunities for future research.

Summary of key findings

Informed by two systematic reviews, the *BMJ Rapid Recommendation* guideline panel[1,2] made recommendations regarding plasma exchange and reduced dose regimen of corticosteroids for patients with ANCA-associated vasculitis. The guideline panel concluded that most (50-90%) of fully informed patients with AAV and with low or low-moderate risk of developing ESKD would decline plasma exchange (weak recommendation), whereas most patients with moderate-high or high risk or requiring dialysis would choose to receive plasma exchange (weak recommendation)[3]. As the reduced dose regimen of glucocorticoids reduces the risk of serious infections and probably does not increase the risk of ESKD, the panel inferred that all or almost all (≥90%) fully informed patients would choose a reduced dose regimen of glucocorticoids (strong recommendation)[3].

Two methodological issues raised in this guideline fueled the development of two GRADE guidance articles (Chapter 3 and 4) and a panel survey approach for eliciting guideline panels' view of patient values and preferences (Chapter 5).

The first GRADE guidance article (Chapter 3) highlighted the importance of clarifying, for every GRADE rating, what it is in which authors rate their certainty of evidence (i.e., the target of certainty of evidence rating)[4]. The guideline provided practical principles for deciding on the target of certainty rating. Authors should consider the degree of contextualization and choose the threshold of interest. The relative location of the point estimate to the chosen threshold determines the target of certainty rating.

When the point estimate is very close to the chosen threshold, GRADE suggested that authors could either still rate their certainty in relation to that chosen threshold and, if the confidence interval crosses the threshold, rate down for imprecision. Alternatively, they could switch to rate certainty in relation to two adjacent thresholds. Using the initial chosen threshold, however, causes counterintuitive problems and might be misleading.

The second GRADE methods article (Chapter 4)[5], focusing on the case of choosing the null effect threshold as the threshold of interest, elaborated on challenges (i.e., rating down for imprecision when the confidence interval is very narrow; rating certainty that there is an effect when the point estimate shows the effect is trivial) and provided solutions. When the threshold is the null and the point estimate is very close to the null, we suggested switching the target of certainty rating from a non-zero effect to a little or no effect. By introducing the concept of range of MIDs this article further discussed how close the point estimate needs to be to the null before authors should consider switching the target of certainty rating from a non-zero effect to a little to no effect.

The report in Chapter 5 introduced a five-step framework for developing and implementing panel surveys in the context of guidelines, and illustrated three different objectives of panel surveys (i.e., establishing an MID threshold; establishing a decision threshold; explicitly specifying the percentage of patients who would elect for or against an intervention)[6]. The users of this framework can choose the objective based on what information regarding patient values and preferences is needed for their guidelines.

The qualitative evaluation in Chapter 6 revealed that most guideline panelists found the surveys were easy to follow and helped guideline panels explicitly consider and incorporate patient values and preferences in making recommendations.

Limitations in methods

In the two GRADE projects, we conducted discussions and presentations within the GRADE Working Group. A first possible limitation is that the approaches we developed may heavily depend on who participated in such discussions and provided feedback. People with different perspectives might come up with diverse approaches to address the same methodological issue. In our projects, the perspectives outside of GRADE were not involved. We do recognize this limitation and acknowledge the approach we suggested might just be one approach among other possible approaches.

Another limitation is that we developed the GRADE approaches mainly through iterative discussions and presentations. People might argue that there is no formal validation study of the approaches (i.e., assessing the validity and reliability of the approaches). Indeed, unlike some other methodological studies, we used real examples in different contexts to evaluate the applicability of the approaches rather than evaluating the content validity, test-retest reliability etc. These usual indicators are often not feasible for assessing GRADE approaches and the applicability in different contexts is probably the most important aspect of GRADE approaches.

Regarding the limitations in methods for developing the panel survey approach, most of our team members are from the GRADE working group. Involving different perspectives outside of GRADE might have benefited the development of the panel survey approach and improved its applicability. However, even for guideline panels who do not apply GRADE, patient values and preferences are probably one of the components they would consider when making recommendations. Our approach for helping elicit panels' inferences regarding how typical patients would trade off the benefits and harms is probably still useful to these other guideline panels.

Finally, people might have concerns regarding guideline panels' ability to design and implement such panel surveys. Indeed, for now, not many guideline panels know this panel survey approach. And for those who know this approach, they might not be able to independently design the survey. We are available for consultation for any guideline panel seeking guidance in creating and implementing panel surveys. We plan to further promote and illustrate this approach through presentations and workshops in future academic conferences.

Further reflections on the two methodological issues

Deciding on the target of GRADE certainty of evidence rating

Explicitly or implicitly setting a threshold

According to current GRADE guidance[4,7], when rating the certainty of evidence, authors need to identify a threshold or range of interest. Authors often ask how to set a threshold other than the null (i.e., the threshold of small, moderate and large effect). Setting these thresholds always involves values and preferences. In Chapter 4 (under *5. Imprecision rating after switching the target of certainty rating*), we introduced some available approaches for explicitly setting thresholds with understanding of patient values and preferences[6,8-11]. In practice, these approaches still have some limitations. For example, due to low methodological quality or small sample size of primary studies, an MID obtained from a systematic review of primary studies might still leave uncertainty regarding the smallest change that patients would perceive as important. Guideline panels might find it difficult to judge whether a threshold based on surveys from patients with different cultures, different beliefs or at different stages of disease other than their target population can be adopted.

Given the difficulty in setting explicit thresholds, authors may consider an implicit approach of gaining insight into a threshold after looking at the effect estimate. Consider authors deciding to rate their certainty of evidence in a large effect (in relation to a large effect threshold). When presented with a particular estimate of effect, they probably find it easier to say whether that effect is or is not a large effect than to specify the threshold that divides a large effect from a moderate effect. In other words, it is reasonable to say the authors are rating their certainty that a large effect exists without, at the outset, specifying the exact threshold that represents the boundary between a large and a moderate effect. This approach also applies to considering the thresholds of a small effect and a moderate effect. For instance, Chapter 3 (Figure 7) shows corticosteroids result in 1.8 fewer deaths per 100 patients, with a 95% CI from 4.1 fewer to 0.8 more deaths per 100 patients. Without specifying an explicit small effect threshold, the authors could consider that 1.8 fewer deaths per 100 patients represents an important effect. Thus, they would rate their certainty that corticosteroids result in an important reduction on death.

Determining the target of certainty of evidence rating in fully contextualized approach

In Chapter 3 and 4, we discussed how to decide the target of certainty of evidence rating considering a single outcome at a time and using the null, small, moderate and large effect thresholds as the threshold or thresholds of interest. What we have not addressed is the decision of target of certainty rating in a fully contextualized approach (i.e., simultaneously considering all outcomes associated with a given decision and the associated value and their relative performance)[7]. Fully contextualized approaches are, however, the final aim of evidence appraisal, particularly when issuing recommendations for clinical practice.

There are some possible approaches for considering the target of certainty in the fully contextualized approach.[12] One option is to operationalize the certainty of evidence ratings both as individual ratings of each outcome and as a single overall rating in net effect (a composite of individual effect estimates of each outcome)[13]. The location of the point estimate of net effect in relation to the null effect determines the target of certainty rating under the fully contextualized

approach. The lowest rating of certainty among the critical individual outcomes generally provides an upper limit for the overall certainty in the net effect[14]. Given the challenges in calculating net effect, however, this option has not been widely used.

Another approach is to identify one key beneficial (or harmful) outcome, and given the harms (or benefits) to set a decision threshold for the minimal benefit (or maximum harm) that the target population would require for accepting the intervention. The relative location of the point estimate of the key beneficial outcome (or the key harmful outcome) to the decision threshold determines the target of certainty rating under the fully contextualized approach. For example, consider a guideline for colorectal cancer screening in adults aged 50-79 years[15]. The panel considered reduction of colorectal cancer related mortality as the key benefit, and inferred that given the harms and burdens of colorectal cancer screening, most people would require a reduction of 10 per 1000 in colorectal cancer related mortality to undergo the procedure. A systematic review revealed that colonoscopy would yield 8 fewer deaths per 1000 adults with a confidence interval from 13 fewer to 5 fewer (Figure 1)[16]. As the point estimate falls below the decision threshold, the target of certainty rating would be colonoscopy has an effect that is smaller than the minimal benefit target population require for accepting colonoscopy.

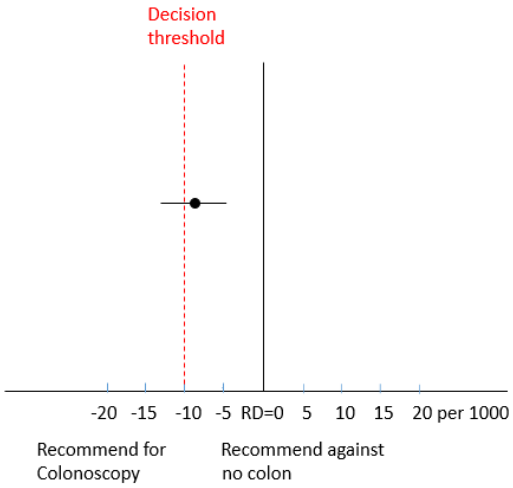


Figure 1 Colonoscopy versus no colonoscopy for adults aged 50-79 years on colorectal cancer related mortality: rating certainty in relation to the decision threshold (10 fewer deaths per 1000 patients), as the point estimate (8 fewer deaths per 1000 patients) the authors would rate certainty that colonoscopy has an effect that is smaller than the minimal benefit the target population require for accepting colonoscopy (i.e., the target of certainty rating).

This second approach has an important limitation: it is only applicable when there is a single key beneficial or harmful outcome. Future research can further explore possible approaches for determining the target of certainty of evidence ratings under fully contextualized approach.

Expressing certainty of evidence using multiple thresholds

Another unsolved issue regarding the target of certainty of evidence rating is: are there any situations in which expressing the certainty of evidence in multiple targets is desirable? If yes, what are these situations?

For example, consider a guideline for preventing clinically important gastrointestinal (GI) bleeding in critically ill patients[17]. The systematic review revealed that, compared with placebo, proton pump inhibitors (PPIs) yielded a reduction of 23 GI bleedings per 1000 patients, with 95% CI from a reduction of 34 to a reduction of 6 per 1000 in patients with high risk of clinical important GI bleeding (Figure 2). The systematic review also revealed that PPIs may increase the risk of pneumonia with low certainty due to very serious imprecision.

Considering the balance between benefit and harm, the certainty of evidence, resources required for implementing the intervention, and feasibility and acceptability of the intervention, the panel could decide: if the PPI has a large effect on reduction of GI bleeding (with moderate or high certainty), the panel would make a strong recommendation for PPI; if the PPI has moderate effect (with any level of certainty), the panel would make a weak recommendation for PPI; if the PPI has small but important effect (with any level of certainty), the panel would make a weak recommendation against PPI; and if the PPI has a little or no effect (with any level of certainty), the panel would make a strong recommendation against PPI.

In such a case, multiple ratings (one for each target) might be desired to inform the direction and strength of recommendation. The first rating would be rating certainty in relation to the large effect threshold. As the point estimate falls below the threshold, the systematic review authors would rate certainty that the true effect is smaller than a large effect. If the authors have no concern on the other four GRADE domains, the certainty of evidence would be high. With such information, the guideline panel would not make a strong recommendation for PPI. Similarly, the authors would express certainty of evidence ratings using the other two thresholds. All these ratings together would inform the guideline panel's final decision on the direction and strength of recommendation.

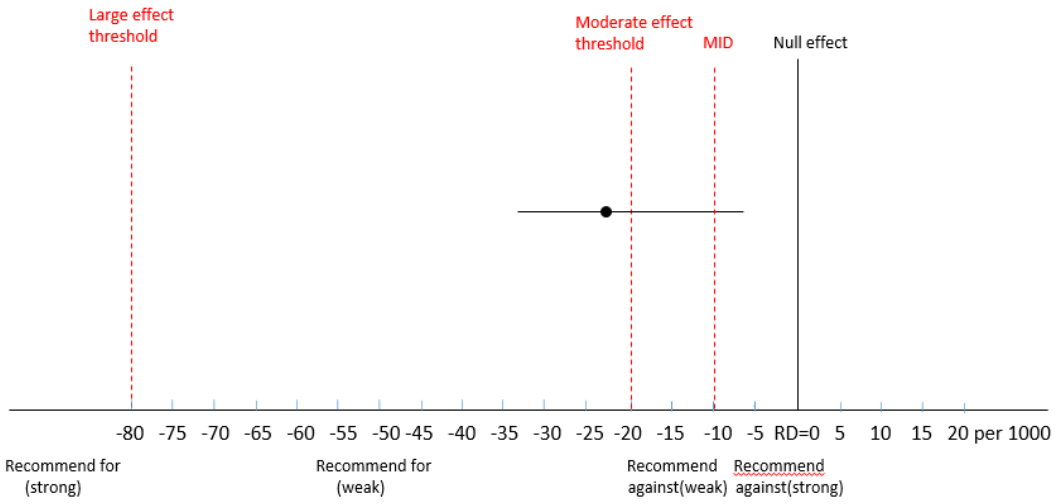


Figure 2 Proton pump inhibitors versus no proton pump inhibitors for patients on gastrointestinal bleeding: assuming no concerns in risk of bias, indirectness, inconsistency and publication bias, there is a high certainty in an effect smaller than large, low certainty in a moderate effect (if rating down two levels for imprecision), moderate certainty in a small but important effect (if rating down

one level for imprecision), high certainty in an effect.

Understanding patient values and preferences, and incorporating into making recommendations

Panel surveys provide a systematic approach for guideline panels interpreting patient values and preferences rather than replace patient surveys

The panel survey approach aims to provide a systematic approach for guideline panels, based on available evidence (e.g., patient survey, focus group commissioned by guideline panel) and panel's experience (e.g., experience in shared decision makings), to estimate patient values and preferences. During the publication peer review process, we received some comments such as "correlation between the panel survey results and patient preferences are needed to more firmly justify your conclusions". Although such comparison, to some extent, can assess guideline panels' ability in making inferences of patient values and preferences, the panel survey should not be treated as a substitution for patient surveys. Panel surveys can also help understand other panel members' understanding regarding patient values and preferences. So no matter whether a patient survey is available or not, guideline panels can apply the panel survey approach facilitating interpretation of patient values and preferences.

Until now, we only applied the panel survey approach in guidelines that take an individual patient perspective. In the future, when applying the approach in guidelines that take other perspectives (e.g., public health perspective), we will further evaluate its applicability and influence on panel's understanding of target population's values and preferences and decisions on recommendations. With further application, we might also be able to expand the objective of the panel survey approach.

Whose values and preferences should be considered in making guideline recommendations

In the qualitative evaluation of the panel surveys (Chapter 6) and a user-testing of educational video of the survey approach[18], a few participants in the interviews (including clinical experts and patient partners) commented that they think both the patients' and the clinicians' values and preferences should be considered.

There might be some situations in which perspectives other than those of the patients should be considered in making guideline recommendations. One situation is the guideline panel might consider externality when making recommendations. For example, a decision on whether to get Polio vaccine not only impacts one's own health but also that of the others (e.g., neighbors, classmates). Thus, should the panel consider the values of the relevant stakeholders in making guideline recommendations?

Another consideration is that values and preferences might be different between those who have experienced the disease or condition (i.e., patients) versus those who are at risk of developing that disease or condition (i.e., non-patients). For example, people who have not had an amputation may tend to place a very high value on the harms and burdens associated with amputation. Half a year after an amputation, however, they may have successfully adapted and place a lower value on the harms and burdens. In such case, should the guideline panel not only consider the values and preferences of the patients but also that of the population at risk?

Indeed, the GIN (Guidelines International Network) - McMaster Guideline Development Checklist, which lists topics and items outlining the practical steps to consider for developing guidelines, suggests that guideline panels should “determine whose perspective(s) will be considered when obtaining information about the relative importance of outcomes and interventions, values, preferences or utilities and when making decisions or formulating recommendations”. The Checklist also lists some possible perspectives including patients, public, society, clinicians[19]. A research team commissioned by the WHO developed a new WHO-INTEGRATE evidence to decision framework aiming to facilitate structured reflection and discussion of guideline development regarding population-level and system-level interventions[20].

This research team argued that the GRADE Evidence to Decision (EtD) framework "does not sufficiently consider the central role of the social and economic determinants of health and implications of health sector or intersectoral interventions for society as a whole" and "the framework may not be entirely suitable to broader public health and health system decision-making contexts". This team further developed a step-by-step guideline for incorporating a complexity perspective in guideline development for public health and health system interventions. They suggested that, in addition to focusing on the direct health benefits or harms associated with the intervention, guideline panels should choose the following criteria for in-depth consideration through evidence collection, synthesis, and assessment: the acceptability of the intervention among different groups of stakeholders, its societal and ecological implications and its impact on health equity, equality and non-discrimination, implying the values and preferences of some others than the patients should be considered. However, this framework does not provide guidance regarding in which situations whose values and preferences should be considered, and how to integrate or weigh different perspectives. We think these issues worth further discussion.

Concluding remarks

This thesis started with the development a *BMJ Rapid recommendation*, and then discussed two methodological issues in the context of making guideline recommendations. We provided suggestions regarding how to decide the target of certainty of evidence rating and provided an innovative approach for incorporating patient values and preferences in making recommendations. Unsolved methodological issues regarding rating certainty of evidence and interpreting magnitude of effect remain providing opportunities for further studies.

1. Walsh M, Collister D, Zeng L, et al. The effects of plasma exchange in patients with ANCA-associated vasculitis: an updated systematic review and meta-analysis. *BMJ*. 2022;376:e064604. doi:10.1136/bmj-2021-064604
2. Xiao Y, Guyatt G, Zeng L, et al. Comparative efficacy and safety of alternative glucocorticoids regimens in patients with ANCA-associated vasculitis: a systematic review. *BMJ Open*. 2022;12(2):e050507. doi:10.1136/bmjopen-2021-050507
3. Zeng L, Walsh M, Guyatt GH, et al. Plasma exchange and glucocorticoid dosing for patients with ANCA-associated vasculitis: a clinical practice guideline. *BMJ*. 2022;376:e064597. doi:10.1136/bmj-2021-064597

4. Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol*. 2022;150:216-224. doi:10.1016/j.jclinepi.2022.07.014
5. Rating certainty when the target threshold is the null, and the point estimate is close to the null (unpublished)
6. Zeng L, Helsingen LM, Bretthauer M, et al. A novel framework for incorporating patient values and preferences in making guideline recommendations: guideline panel surveys. *J Clin Epidemiol*. 2023;S0895-4356(23)00175-0. doi:10.1016/j.jclinepi.2023.07.003
7. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006
8. Helsingen LM, Zeng L, Siemieniuk RA, et al. Establishing thresholds for important benefits considering the harms of screening interventions. *BMJ Open*. 2020;10(12):e037854. doi:10.1136/bmjopen-2020-037854
9. Neumann I, Quiñelen E, Nahuelhual P, et al. Using Explicit Thresholds were valuable for judging Benefits and Harms in partially contextualized GRADE Guidelines. *J Clin Epidemiol*. 2022;147:69-75. doi:10.1016/j.jclinepi.2022.03.017
10. Devji T, Carrasco-Labra A, Guyatt G. Mind the methods of determining minimal important differences: three critical issues to consider. *Evid Based Ment Health*. 2021;24(2):77-81. doi:10.1136/ebmental-2020-300164
11. PROMID. https://www.promid.org/
12. Schünemann H, Brożek J, Guyatt G, et al. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. 2013. https://gdt.gradepro.org/app/handbook/handbook.html
13. Alper BS, Oettgen P, Kunnamo I, et al. Defining certainty of net benefit: a GRADE concept paper. *BMJ Open*. 2019;9(6):e027445. doi:10.1136/bmjopen-2018-027445
14. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66(2):151-157. doi:10.1016/j.jclinepi.2012.01.006
15. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;367:l5515. doi:10.1136/bmj.l5515
16. Jodal HC, Helsingen LM, Anderson JC, Lytvyn L, Vandvik PO, Emilsson L. Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. *BMJ Open*. 2019;9(10):e032773. doi:10.1136/bmjopen-2019-032773
17. Ye Z, Reintam Blaser A, Lytvyn L, et al. Gastrointestinal bleeding prophylaxis for critically ill patients: a clinical practice guideline. *BMJ*. 2020;368:l6722. doi:10.1136/bmj.l6722
18. User testing of an educational video for introducing panel survey approach (unpublished)
19. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ*. 2014;186(3):E123-E142. doi:10.1503/cmaj.131237
20. Rehfuess EA, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. *BMJ Glob Health*. 2019;4(Suppl 1):e000844. doi:10.1136/bmjgh-2018-000844