PANEL ENGAGEMENT IN HEALTH GUIDELINE DEVELOPMENT

PRIORITY TOPICS FOR PANEL ENGAGEMENT IN HEALTH GUIDELINE DEVELOPMENT

By WOJTEK WIERCIOCH, BHSc, MSc

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

McMaster University DOCTOR OF PHILOSOPHY (2020) Hamilton, Ontario (Health Research Methodology).

TITLE: Priority Topics for Panel Engagement in Health Guideline Development

AUTHOR: Wojtek Wiercioch, BHSc, MSc (McMaster University)

SUPERVISOR: Professor Holger J. Schünemann, M.D., Ph.D., M.Sc.

NUMBER OF PAGES: xx, 159

LAY ABSTRACT

Health care guidelines, also referred to as clinical practice or public health guidelines, involve summarizing the available research evidence on a given health care topic and issuing recommendations about the best care. Guidelines allow clinicians, patients, health administrators and policy-makers to be efficiently informed and stay up to date on alternative care options, such as the best current treatments and strategies to diagnose various diseases and health conditions. Developing a guideline is a complex and multidisciplinary process that includes involving a panel of experts, typically consisting of clinicians, patients, public health professionals and other providers or consumers of health care. The panel is involved in selecting the health care questions to address (e.g. specific treatments or diagnostic strategies to evaluate), reviewing a summary of the evidence from research studies, and making judgements about benefits and harms of alternate options or strategies. The panel then formulates recommendations that give guidance on what the best options are to use for the health condition in question. The steps and approaches to develop a guideline that is considered trustworthy have been established over the past decades, including universally accepted standards. However, there remain critical research questions on how to best reach these standards, including how to best engage guideline panels in the steps.

The research work presented in this thesis focuses on proposing and evaluating new methods and approaches for guidelines panels to make decisions about health

outcomes, priority health care questions for guidelines, and to evaluate the guideline development process. It includes three studies: 1) a study on creating health outcome descriptors with panels to provide a commonly accepted definition of a health outcome; 2) a study to evaluate specific criteria that panels can use to prioritize health care questions, and an approach to judge the importance of health outcomes; and 3) a study to develop a survey instrument for guideline panel members to evaluate the guideline development process they participate in.

In these studies, we established an approach for creating the health outcome descriptors with panels, which helped in keeping consistency with how panels understood and considered different health outcomes throughout the guideline development process. The criteria evaluated for prioritizing healthcare questions informed panel discussions and selection of questions for their guideline topics. The proposed approach for judging the importance of health outcomes helped panels to select what the critical outcomes were for making decisions about the benefits and harms of alternate options or strategies. Finally, the survey tool we created allowed members of guideline panels to provide feedback on strengths and weaknesses and areas for improvement in the process after they participated in developing a guideline. Our findings will allow organizations responsible for guideline development to apply the new methods with their panels and to evaluate their guideline processes to inform quality-improvement efforts.

ABSTRACT

Health care guidelines provide a means of assessing the best available research evidence on a given health care topic and offering recommendations about use of specific interventions and management of patient care. Guidelines allow clinicians, patients, health administrators and policy-makers to be efficiently informed and stay up to date on alternative care options. The development of guidelines is a complex and multidisciplinary process, with a defining feature of involving a panel of experts in steps such as selecting health care questions, assessing the research evidence, making judgements about health benefits and harms, and, ultimately, formulating recommendations. Guideline methodology has advanced over the past decades, including establishment of specific steps and standards to ensure trustworthiness of guidelines. However, there remain critical research questions on how to best accomplish and reach these standards, including how to best engage panels in the steps.

This thesis presents a body of research on the development and evaluation of new methods for decision-making and considering health outcomes in guidelines, prioritizing health care questions for guidelines, and evaluating the guideline development process. It includes three studies: 1) a methodological study on developing health outcome descriptors to define health outcomes considered in decision-making by guideline panels; 2) a methodological study and randomized controlled trial to evaluate specific criteria for panels to consider when prioritizing health care questions for guidelines and to judge the importance of health outcomes; and 3) an instrument

development and validation study to create a tool for panel members to evaluate the appropriateness of the guideline development process they participate in.

In these studies, we established a method and steps for creating health outcome descriptors with panels, aimed at achieving consistency in how health outcomes are considered throughout the guideline development process, from prioritization to formulating a recommendation on the basis of those outcomes. The structured approach and criteria evaluated for prioritization of healthcare questions informed panel deliberations and decisions about choosing questions for their guideline topics, and the proposed methods for outcome prioritization facilitated panels in informing what the critical and important outcomes were for decision-making. Finally, the instrument we developed facilitated members of guideline panels to provide their assessment of the guideline development process and identify strengths and weaknesses and areas for improvement. Our findings will allow organizations responsible for guideline development to apply the new methods with their panels and to evaluate their guideline processes to inform quality-improvement efforts.

ACKNOWLEDGEMENTS

I wish to express my gratitude to those who have helped me through this journey, enhancing my experiences and learning.

I would like to express my deepest gratitude to my supervisor, Dr. Holger Schünemann, for his continuous support to strive for professional and personal growth. Thank you for your encouragement and leadership that offered countless opportunities for a rich learning and research environment. I would also like to express my gratitude to the members of my thesis committee: Dr. Elie Akl, Dr. Meghan McConnell, and Dr. Nancy Santesso, who have played a significant part of this journey, offering their enthusiasm, encouragement and direction.

Thank you to the staff at the department of Health Research Methods, Evidence, and Impact (HEI) for helping me along the way. A special thank you to my friends and peers at HEI for your support, and the enjoyment and motivation you brought into the journey.

Finally, I want to dedicate this achievement to my family for their endless support and encouragement.

Wojtek Wiercioch

Hamilton, Ontario – September 2020

PREFACE

The work in this dissertation is presented as three manuscripts which have been accepted for publication or prepared for submission framed by an introduction (Chapter 1) and summary and conclusion (Chapter 5). The manuscript in Chapter 2, "Development and application of health outcome descriptors in the production of practice guidelines", is prepared for submission to the Journal of Clinical Epidemiology. The manuscript in Chapter 3, "New methods facilitated the process of prioritizing questions and health outcomes in guideline development", will be submitted to the Journal of Clinical Epidemiology. These manuscripts in Chapters 2 and 3 are intended to be submitted as companion papers in a two-part series. The manuscript in Chapter 4, "Assessing the process and outcome of the development of practice guidelines and recommendations: PANEVIEW instrument development", has been accepted in a general medical journal, the Canadian Medical Association Journal.

The work presented in Chapter 2 served to support guideline development with the American Society of Hematology. I actively contributed to the development of the protocol for the research, production of health outcome descriptors, and drafted the manuscript which is circulated to co-authors. I incorporated feedback from co-authors, prepared and finalized it for submission. Chapter 3 was a methods study that I conceived and coordinated with supervisor Dr. Holger Schünemann. I developed the online surveys, managed the recruitment of participants, completed the analysis and prepared the manuscript for submission after integration of comments from all co-authors.

Chapter 4 describes the development of an instrument and tool that I conceived of with my supervisor Dr. Holger Schünemann and Dr. Elie Akl. I coordinated the work, performed major aspects of the research as described in the manuscript along with a research team. I coordinated the administration of the instrument and collection of data and Dr. Meghan McConnell and I conducted the analysis. I drafted the manuscript, incorporated feedback from the co-authors and prepared the manuscript for submission. The work was funded by the American Society of Hematology and the McMaster GRADE Centre.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

LIST OF APPENDICES

Chapter 2

Chapter 3

Chapter 4

LIST OF ABBREVIATIONS

AGREE - Appraisal of Guidelines for Research and Evaluation

ASH – American Society of Hematology

CI - Confidence Interval

COI - Conflict of Interest

DVT – Deep vein thrombosis

EtD – Evidence-to-decision

GIN – Guidelines International Network

GRADE – Grading of Recommendations Assessment, Development and Evaluation

HOD – Health outcome descriptor

PE – Pulmonary embolism

PICO - population, intervention, comparator, outcomes

RIGHT - Reporting Tool for Practice Guidelines in Health Care

SD – Standard deviation

SoF – Summary-of-findings

VAS – Visual analogue scale

VTE – Venous Thromboembolism

DECLARATION OF ACADEMIC ACHIEVEMENT

I declare that I, jointly with my supervisor, Professor Holger J. Schünemann, played the primary role in the conception, design, and execution of the studies here included. We obtained feedback and advice from Professors Akl, McConnell and Santesso, as well as from clinical and methodological experts that co-authored the work.

This work is original research that I conducted. I am the principal contributor and first author of all the manuscripts contained in this dissertation. I am responsible and made the following contributions to all projects included in this dissertation: conception, study design, data acquisition, analysis, and writing of materials. I contributed to the development of the protocol, production of health outcome descriptors, and completed the data analysis for the methodological study on use of health outcome descriptors in guideline development (Chapter 2). I designed surveys, coordinated data collection from guideline panels, and managed and analyzed the data for the methodological study on question prioritization and rating of outcomes in guideline development (Chapter 3). For the instrument development study (Chapter 4), I conducted the systematic review, including performing searches, screening, and data abstraction. I coordinated and participated in research team consensus meetings. I designed surveys and interview guides, and collected and analyzed data.

I conducted analyses, designed figures and tables, and organized research team meetings. I wrote the manuscripts with editorial advice and supervision of Professor Schünemann, and with feedback from Professors Akl, McConnell and Santesso. The

authors on each paper contributed significantly with important comments and advice for the final manuscripts.

For all three manuscripts in this "sandwich" thesis, earlier parts of this research work have been presented at international academic conferences. The first (Chapter 2) and second manuscripts (Chapter 3) will be submitted to the *Journal of Clinical Epidemiology*. The third manuscript (Chapter 4) was accepted for publication in the *Canadian Medical Association Journal* in June 2020.

CHAPTER 1. INTRODUCTION

Health Care Guidelines

Health care guidelines are intended to inform management of patient care and optimal choice of health care options. Guidelines are defined by the Health and Medicine Division of the National Academies of Sciences, Engineering, and Medicine (formerly the Institute of Medicine) as "statements that include recommendations, intended to optimize patient care, that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options."[1] The definition outlines two key components of a guideline. First, guidelines are expected to provide a systematic summary of all available evidence for a given question or healthcare problem. Second, they are expected to provide clear, actionable statements to give guidance to target users. Given their purpose, health care guidelines are of interest to national organizations, health care professional societies, care providers, ministries of health, policy makers, patients and the public. With broad application and use to inform healthcare decisions, appropriate development of guidelines is essential to ensure their credibility and trustworthiness.

Approaches to Development of Guidelines

Guideline development methods and processes have been outlined, formalized, and evolved over time. Developing a guideline requires specialized knowledge with participation from several groups or stakeholders, including a group to provide oversight, a working group responsible for technical aspects of evidence synthesis, and a guideline panel tasked with reviewing evidence and formulating recommendations. In

early work, the Institute of Medicine defined guidelines as "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances",[2] lacking the explicit description of the systematic review component that is present in the updated definition described above. Systematically developed statements could consist of guideline recommendations simply based on guideline panel experts that did not explicitly describe evidence supporting consensus, and early systems of classifying levels of evidence indeed included expert opinion as a separate level of evidence.[3,4] With the recognition of the need for systematic reviews to underpin guidelines, various systems were developed for evaluating the quality of the evidence in systematic reviews on which recommendations are based.[5] The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach sought to unify and expand criteria that were proposed by earlier systems that considered necessary for determining the quality of evidence.[6] It became widely adopted by guideline developing organizations.[7] While methods to present summaries of evidence and its quality aim to address the component of evidence synthesis in guideline development, recommendation statements that suggest a given action or choice of health care option are the second key component of a guideline. One recent study found that given the option of being presented evidence summaries alone or recommendations accompanying the evidence summaries, in particular when evidence is sparse, approximately 80% of participating clinicians preferred to have recommendations.[8]

Given this need, advancements in methods and development of formal systems for moving from evidence to recommendations and assessing their strength have followed the advancements in evidence synthesis and judging quality or certainty of evidence.[9] Within the GRADE approach this has included the development and application of Evidence-to-Decision (EtD) frameworks to support transparent decision-making by guideline panels when formulating recommendations.[10,11] In order to formulate appropriate recommendations, panels need to adequately consider the synthesized evidence to judge the health benefits and harms of health care options, along with other factors such as patients' values and preferences, costs, impact on health equity and feasibility. The development of these methods and processes has enabled the evolution of guidelines to be evidence-based. However, this has not precluded the need for expert involvement and expert judgement in guideline development.

Formalizing the methods has demonstrated the complexity of the task of guideline development, and the need to have appropriate steps and processes in place to arrive at trustworthy recommendations. These include steps such as appropriate priority setting of guideline topics and questions, having the right group composition for the guideline panel and participating experts, considering the appropriate health outcomes, synthesizing the evidence for those outcomes, having a process to reach consensus on recommendations, adequate reporting, maintaining a well-functioning group process, and managing conflicts of interest. Various methods manuals,

publications, and guidelines for guidelines have summarized the steps,[12-20] and moreover, international standards for guideline development have been established incorporating these steps and processes.[1,21] Among these steps in guideline development, there are a number of priority areas for involvement of the guideline panel.

Prioritization and Decision-Making in Guidelines

A key challenge in ensuring appropriate development and relevance of a health care guideline is that of prioritization to get the work done. Guideline developers and guideline panels face this challenge at several steps, including deciding on the overall guideline topic and specific healthcare questions to address.[16,22] Given resource constraints and the need for timeliness in issuing guidance on a specific health topic, guideline panels must select questions that the guideline will address and for which they will formulate recommendations that are informative for the target audience. How to best involve the guideline panel and conduct the prioritization exercise in guidelines remains an area of need for research.[23]

To formulate recommendations for the healthcare questions prioritized to be addressed in a guideline, the guideline panel will need to weigh the health benefits and harms of alternative care options by assessing the impact of those options on specific health outcomes.[19,22] This, as well, requires prioritization of the health outcomes on which to base the recommendation. Research evidence will be synthesized for these specific outcomes and the guideline panel will deliberate about the balance of health

benefits and harms on the basis of the outcomes. Throughout the process, members of the guideline panel must also be able to maintain a common understanding of the health outcomes that are being considered. The appropriate selection of health outcomes and consideration of the outcomes in decision-making is another priority area for panel engagement.

Evaluation of Guideline Development

As health care guidelines have the potential to influence practice and impact a large number of health care decisions, their rigour and trustworthiness must be ensured and quality improvement in guideline development is an important area of focus. Trustworthy guideline production requires coordination of the steps including administration and organization of the panel and its activities, evidence review, training, and facilitation of panel meetings.[20] Various group processes are also involved in the deliberations by the guideline group, in synthesizing evidence, and reaching consensus to issue recommendations.[14,24] Problems in the guideline process may occur due to issues such as influence of individuals with strong opinions or conflicts of interest, inappropriate panel composition, or recommendations not being consistent with the best available evidence.[25-28] Available instruments to assess the trustworthiness of health care guidelines involve the evaluation of the guideline report, which may not capture the evaluation of the guideline process as it occurred.[29,30] Panel members may provide insight into the process and steps that take place and the evaluation of guideline development is another priority area for panel engagement.

Goals and Scope

This dissertation highlights priority areas for guideline panel engagement with a focus on development and evaluation of new methods in guideline development and the following goals:

1) To describe the methods for development of health outcome descriptors intended to facilitate the prioritization and consideration of health outcomes in guidelines, and the application of the approach in a guideline development project.
2) To describe the development and evaluation of new approaches for prioritizing questions and health outcomes in the context of a guideline development project.
3) To develop and validate a universal tool for a participating guideline panel to assess guideline processes, methods and outcomes of guideline development.

With these goals, the overall objective of the research work presented is to introduce methods, tools, and approaches for guideline development and panel engagement that address current areas of research need. The methods and approaches may be used to inform future guideline development and guideline quality improvement efforts.

Thesis Overview

This dissertation consists of three research sections, described in Chapters 2 to 4, and a fourth section in Chapter 5 providing a summary of the research, conclusions and implications for future research and guideline development.

Chapter 2 presents a methodology for developing and using health outcome descriptors to facilitate the prioritization of health outcomes and decision-making about health outcomes in the context of guideline development. We describe the approach and its application with a group of methodologists, clinicians and panel members in a large-scale guideline development project on the management of venous thromboembolism (VTE) with the American Society of Hematology (ASH). Chapter 3 describes a new method for prioritizing healthcare questions with guideline panels participating in the ASH VTE guidelines project and evaluation of the approach in a randomized controlled trial (RCT). It also describes a methodology for rating the importance and utility of health outcomes with guideline panels. Chapter 4 introduces the PANELVIEW instrument, which allows guideline developers to obtain an evaluation of their processes and methods from the perspective of the participating guideline panel members. We describe its development, informed by a systematic review of the literature and formal item generation and item reduction steps, and results of the application of the instrument with guideline panels in field testing. In concluding remarks, Chapter 5 presents a summary of the research work and direction on how guideline developers may apply the methods to engage their panels in their guideline

development and quality improvement efforts, as well as remaining challenges and opportunities for future research.

References

1. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: 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. Institute of Medicine Committee to Advise the Public Health Service on Clinical Practice G. In: Field MJ, Lohr KN, eds. *Clinical Practice Guidelines: Directions for a New Program.* Washington (DC): National Academies Press (US), Copyright (c) National Academy of Sciences.; 1990.
3. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Canadian Medical Association journal.* 1979;121(9):1193-1254.
4. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. *BMJ (Clinical research ed).* 1999;318(7183):593-596.
5. Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC health services research.* 2004;4(1):38-38.
6. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ (Clinical research ed).* 2004;328(7454):1490-1490.
7. 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.
8. Neumann I, Alonso-Coello P, Vandvik PO, et al. Do clinicians want recommendations? A multicenter study comparing evidence summaries with and without GRADE recommendations. *Journal of clinical epidemiology.* 2018;99:33-40.
9. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *BMJ (Clinical research ed).* 2008;336(7652):1049-1051.
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.
11. Alonso-Coello P, Schunemann 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.
12. Boyd E, Bero L. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. *Health Research Policy and Systems.* 2006;4(1):16.
13. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. *Health Research Policy and Systems.* 2006;4(1):15.

14. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 5. Group processes. *Health Research Policy and Systems.* 2006;4(1):17.
15. Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 14. Reporting guidelines. *Health Research Policy and Systems.* 2006;4:26.
16. Eccles MP, Grimshaw JM, Shekelle P, Schünemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implementation Science.* 2012;7(1):60.
17. Schunemann HJ, Woodhead M, Anzueto A, et al. A guide to guidelines for professional societies and other developers of recommendations: introduction to integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proceedings of the American Thoracic Society.* 2012;9(5):215-218.
18. Shekelle P, Woolf S, Grimshaw JM, Schünemann HJ, Eccles MP. Developing clinical practice guidelines: reviewing, reporting, and publishing guidelines; updating guidelines; and the emerging issues of enhancing guideline implementability and accounting for comorbid conditions in guideline development. *Implementation Science.* 2012;7(1):62.
19. Woolf S, Schünemann HJ, Eccles MP, Grimshaw JM, Shekelle P. Developing clinical practice guidelines: types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations. *Implementation Science.* 2012;7(1):61.
20. Schünemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *Canadian Medical Association Journal.* 2014;186(3):E123-E142.
21. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med.* 2012;156(7):525-531.
22. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of clinical epidemiology.* 2011;64(4):395-400.
23. El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. *BMC Health Services Research.* 2019;19(1):692.
24. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. *Journal of Evaluation in Clinical Practice.* 2001;7(2):165-173.

25. Pagliari C, Grimshaw J. Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. *J Eval Clin Pract.* 2002;8(2):145-153.
26. Hopthrow T, Feder G, Michie S. The role of group decision making processes in the creation of clinical guidelines. *Int Rev Psychiatry.* 2011;23(4):358-364.
27. Neumann I, Karl R, Rajpal A, Akl EA, Guyatt GH. Experiences with a novel policy for managing conflicts of interest of guideline developers: a descriptive qualitative study. *Chest.* 2013;144(2):398-404.
28. Alexander PE, Li SA, Gionfriddo MR, et al. Senior GRADE methodologists encounter challenges as part of WHO guideline development panels: an inductive content analysis. *Journal of clinical epidemiology.* 2016;70:123-128.
29. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *Canadian Medical Association Journal* 2010;182(18):E839-842.
30. Chen Y, Yang K, Marušić A, et al. A reporting tool for practice guidelines in health care: The right statement. *Annals of Internal Medicine.* 2017;166(2):128-132.

CHAPTER 2. DEVELOPMENT AND APPLICATION OF HEALTH OUTCOME DESCRIPTORS IN THE PRODUCTION OF PRACTICE GUIDELINES

Development and Application of Health Outcome Descriptors in the Production of Practice Guidelines

Authors

Wojtek Wiercioch[1,2]
Robby Nieuwlaat[1,2]
Philipp Dahm[3,4]
Alfonso Iorio[2,5]
Reem A. Mustafa[2,6]
Ignacio Neumann[7]
Bram Rochwerg[2,5]
Veena Manja[8,9]
Pablo Alonso-Coello[10]
Thomas L. Ortel[11]
Nancy Santesso[1,2]
Sara K.Vesely[12]
Elie A. Akl[2,13]
Holger J. Schünemann[1,2,5]

for the *ASH VTE Panels Working Group:*

Shannon M. Bates[5]
Adam Cuker[14]
Wendy Lim[3]
Paul Monagle[15]
Rob Kunkle[16]
Daniel M. Witt[17]
Susan R. Kahn[18]
Claire McLintock[19]
Suely M. Rezende[20]
Neil A. Zakai[21]

Author Affiliations:

1. Michael G. DeGroote Cochrane Canada & McMaster GRADE Centres, McMaster University, Hamilton, ON, Canada
2. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
3. Urology Section, Minneapolis VA Health Care System, Minneapolis, MN, USA
4. Department of Urology, University of Minnesota, Minneapolis, MN, USA
5. Department of Medicine, McMaster University, Hamilton, ON, Canada
6. Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical
7. Department of Internal Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile

8. Department of Surgery, University of California Davis, CA, USA
9. Department of Medicine, Veterans Affairs Northern California Health Care System, Mather, CA
10. Iberoamerican Cochrane Center, Biomedical Research Institute Sant Pau (IIB Sant Pau-CIBERESP), Barcelona, Spain
11. Departments of Medicine and Pathology, Duke University Medical Center, Durham, NC
12. Department of Biostatistics and Epidemiology, Hudson College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
13. Department of Internal Medicine, American University of Beirut, Lebanon
14. Department of Medicine and Department of Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
15. Department of Paediatrics, University of Melbourne, Department of Haematology, Royal Children's Hospital, Haematology Research, Murdoch Children's Research Institute, Melbourne, Australia
16. American Society of Hematology, Washington, DC
17. Department of Pharmacotherapy, College of Pharmacy, University of Utah, Salt Lake City, UT, USA
18. Department of Medicine, McGill University and Lady Davis Institute, Montreal, QC Canada
19. National Women's Health, Auckland City Hospital, Auckland, New Zealand
20. Department of Internal Medicine, Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
21. University of Vermont Larner College of Medicine and the University of Vermont Medical Center. Departments of Medicine and Pathology & Laboratory Medicine. Burlington, VT

Corresponding author:
Holger J. Schünemann
Department of Health Research Methods, Evidence and Impact
McMaster University, HSC-2C
1280 Main St West; Hamilton, ON, L8N 3Z5, Canada
e-mail: schuneh@mcmaster.ca

Abstract

Background: Health-related outcomes are the basis by which clinicians, patients, guideline developers and other stakeholders balance the potential benefits and harms of interventions and treatments. In this paper we report on the development of health outcome descriptors (HODs) as a tool to inform the development of guideline recommendations.

Methods: For ten guidelines on the management of venous thromboembolism, pairs of panelists with topic expertise drafted HODs using an instructional guide and template prepared by a group of guideline methodologists. We used the HODs in outcome prioritization and utility rating exercises, and in the process of formulating recommendations.

Results: We developed 127 health outcome descriptors for 104 health outcomes across guideline topics. HODs described the symptoms, time horizon, testing and treatment, and consequences associated with the health outcomes. There was 82% agreement in the categorization of outcome importance across the ten guideline panels, and outcome importance ratings were strongly associated with panelists' utility ratings of the outcomes. HODs for prioritized outcomes facilitated panel deliberations when incorporated into Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence-to-decision frameworks.

Conclusions: The use of HODs is intended to create common definitions of outcomes considered across multiple steps of guideline development and linked across different stakeholders within a given clinical or health policy topic area.

Keywords: clinical practice guidelines, health outcomes, clinical decision-making, health recommendations, GRADE

What is new:

- Health outcome descriptors (HODs) seek to provide a unified definition of health outcomes, focusing on consequences for people experiencing those outcomes.
- The approach can be applied by guideline developers to help avoid differences in implicit understanding of health outcomes by panel members and other stakeholders across the steps of guideline development.
- The structured approach of defining health outcomes with HODs to inform formulation of recommendations also aims to facilitate communication of outcomes across clinicians, guideline developers, researchers and policy-makers.

Ph.D. Thesis – Wojtek Wiercioch; McMaster University – Health Research Methods, Evidence, and Impact.

Abbreviations

ASH – American Society of Hematology
DVT – Deep vein thrombosis
EtD – Evidence-to-decision
GIN – Guidelines International Network
GRADE – Grading of Recommendations Assessment, Development and Evaluation
HOD – Health outcome descriptor
PE – Pulmonary embolism
SD – Standard deviation
VAS – Visual analogue scale
VTE – Venous Thromboembolism

1. Introduction

Health-related outcomes are the basis by which clinicians, patients, guideline developers and other stakeholders balance the potential benefits and harms of interventions and treatments. While approaches such as defining core outcome sets provide guidance on what outcomes are important in a specific clinical research area, the definitions of outcomes can vary and lead to inconsistent understanding by different stakeholders.[1] For example, patient-important outcomes may differ according to their location (e.g., proximal or distal deep vein thrombosis (DVT)), presentation (symptomatic or asymptomatic DVT), or severity (e.g. a mild or severe occurrence of DVT). Consequently, how these specific outcomes will be managed by the treating clinicians and how they will be perceived by those experiencing them will differ. When making healthcare decisions informed by research evidence, ensuring that there is a common understanding of the health outcomes presented is, thus, critical to determining the balance of effects of an intervention and deciding to use it.

Guideline panels are faced with these challenges in several ways. First, panels must decide what are the important health outcomes on which to base their recommendations.[2,3] Second, panels supported by evidence synthesis experts, must identify the research evidence that describes the effects of interventions for those important outcomes. Third, panels must deliberate and discuss the balance of benefits and harms of interventions on the basis of those health outcomes.[4,5] Lastly, panels must communicate the recommendation and how they arrived at it to the target audience;

clinicians, patients, and other stakeholders involved in making healthcare decisions. This justification to the guideline user involves describing the impact of the recommended intervention on health outcomes in the target population. In each of the steps, maintaining a common understanding of the health outcomes that are being considered is integral to the process of arriving at a trustworthy recommendation and to avoid problems arising from different implicit definitions of health outcomes amongst the various guideline stakeholders.

One approach to overcome these challenges may be to formally define health outcomes that are considered across the various steps of guideline development, with active involvement of guideline panels. Health outcome descriptors (HODs) (also referred to as 'health marker states') are a new approach using a structured format to define a health outcome, focusing on the persons experiencing the outcome and the implications for managing the outcome. In this article, we describe methods for developing HODs in the production of practice guidelines and results of how the approach worked in a guideline development project when applying it to prioritizing and rating health outcomes, as well as for decision-making by guideline panels. In a companion publication we report on the detailed methods on how the outcome prioritization and rating exercises were conducted in the same guideline development project.[6]

2. Methods

In a recent guideline development project, the American Society of Hematology (ASH) set out to develop ten guidelines on management of venous thromboembolism (VTE).[7-13] The guideline development was conducted in collaboration with the McMaster University GRADE Centre and the general methods followed the Guideline International Network (GIN)-McMaster Guideline Development Checklist and the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach,[14] which are described in detail elsewhere.[15] As part of the process we formed a methods advisory group consisting of methodology and clinical experts who met regularly during the project to advise and decided about specific methods and approaches to use to produce the guidelines. This included advising on and planning the approach for developing and using HODs. For each of the guidelines, ASH formed a multidisciplinary panel consisting of clinical and methodology experts and patient representatives, with each panel led by a clinical chair and methodology co-chair. The methods advisory group included all methodology chairs and some of the clinical chairs who were able to participate in weekly preparatory meetings.

To begin the development of their guidelines, the panels first prioritized healthcare questions to answer through recommendations, and participated in a health outcome brainstorming exercise to identify potentially critical or important health outcomes to consider when formulating the recommendations.[2] Panels then participated in an outcome prioritization exercise to select the final outcomes, followed

by an outcome utility rating exercise.[6,15] To prepare for these initial steps and potential challenges with inconsistent perceptions of health outcomes, we established an approach and developed HODs for all outcomes identified by panels in the brainstorming step, which would then be applied in the subsequent steps.

Development of health outcome descriptors

We iteratively developed a protocol and instructional guide for the development of HODs by panel members, which included a HOD template and examples. Figure 1 shows the HOD template and the instructional guide is provided in Appendix A. The basis for the guide was previous work, including valuation of health outcomes by patients with pulmonary or gastrointestinal disease.[16-19] The methods advisory group reviewed and provided feedback to finalize the guide, and planned the steps for applying the approach over multiple teleconferences.

The HOD template presented in the guide included four domains to describe the symptoms, time horizon, testing and treatment, and consequences associated with the health outcome. The domains were described using bullet points for ease of reading and conciseness.[18] Across the domains, the intended focus for the HOD was to cover the essential characteristics and usual representation of the health outcome from the perspective of the patient or healthcare recipient. If the characteristics would be considered to vary for a health outcome (e.g. pain, or allergy), then separate HODs were necessitated (e.g. acute and chronic pain, or mild allergy and severe allergy). The guide

instructed HOD developers to write descriptions at a Flesch-Kincaid Grade 9 readability level, using active language ("you").

Because the guideline panels would participate in an outcome importance rating exercise to prioritize outcomes viewed as critical for decision-making as well as an outcome utility rating exercise, which we describe below, we prepared two HOD templates; one for importance rating and one for utility rating (see Appendix A). Both templates had the same structure and domains, but the template for utility rating was intended to expand on the outcome importance rating with additional details (e.g. two to four bullet points per domain rather than one or two points) that would allow for more specificity when eliciting a utility value on a 0 to 100 visual analogue scale (VAS), where a value of 0 indicates the state of being dead and 100 indicates the state of full health.[16]

Participants and Coordination

We sought participation from the panel clinical chairs, methodology chairs, and panel members, with experience in the respective clinical areas to draft the HODs according to the template. We created an online sign-up list, compiling all of the outcomes from panels' initial brainstorming, for clinical chairs and methodologists to volunteer to draft HODs and request participation from members of their panel, particularly if they considered specific panel members as having the best expertise to draft an HOD for certain outcomes, as well as to split the workload. Participants were

provided the instructional guide and we addressed questions arising during HOD development in our regular methods advisory group meetings.

Following the detailed guide and examples, the panelists were instructed to consider available sources of information for drafting the HODs, which included any known existing health outcome descriptors, items from health-related quality of life measurement tools, and their clinical observations and expert experience. A panelist or methodologist drafted each HOD, with review by at least one clinician expert. All HODs were then reviewed by another methodologist (HJS or RN) to ensure consistency in the terminology and language used, and ensure adherence to the template and instructional guide.

Outcome prioritization, outcome utility rating, and evidence-to-decision process

Once HODs were finalized, we prepared outcome prioritization and utility rating surveys for panelists to complete individually. Detailed methods for these two steps are reported in the companion publication.[6] Briefly, panelists rated outcomes from the initial brainstorming, which were now described with the HOD, using a 9-point scale according to the GRADE approach, with a rating of 7 to 9 indicating the outcome as being critical for decision-making, 4 to 6 as important but not critical, and 1 to 3 of limited or no importance.[2] These ratings informed panels' discussions and selection of the final outcomes to consider for their guideline questions.

In a subsequent step, which took place towards the end of the guideline process prior to final panel meetings to formulate recommendations, panelists rated the health

utility of outcomes on a VAS of 0 (representing the state of being dead) to 100 (representing the state of full health), considering the perspective of the person experiencing that outcome. The utility rating exercise was intended to elicit information about from panellists that would be used in the evidence-to-decision step to help panels consider the relative value attached to outcomes.[6,15] We hypothesized that using the HODs would allow for a common understanding of outcomes by panelists and asked the ten panels to rate the importance and utility for the same, complete list of outcomes. Panelists were, therefore, asked to rate outcomes also outside of their area of expertise (e.g. panelists of guideline on VTE prevention in surgical patients would rate outcomes identified by panelists of guideline on pediatric VTE).

We used GRADE evidence-to-decision (EtD) frameworks to facilitate formulation of recommendations during panel meetings.[4,5] The frameworks included Summary-of-Findings tables presenting research evidence from systematic reviews and estimates of effects of interventions for the prioritized outcomes.[20,21] They were encouraged to refer to the HODs when reviewing the effects on outcomes in the SoF tables. To reach consensus on the direction and strength of recommendations, panel members made judgements for evidence-to-decision criteria including the magnitude of benefits and harms, balance of benefits and harms, patients' values and preferences, resource use and cost-effectiveness, impact on health equity, feasibility, and acceptability.

Analysis

For descriptive analysis of outcome importance and utility ratings we calculated means and standard deviations. To classify the outcome importance ratings into the 3 categories of the 9-point scale we rounded the mean ratings to the nearest full digit. We also calculated agreement of the categorization of outcome importance across the ten guideline panels. For the purpose of analysis, we transformed the utility ratings from the VAS to a 0 (dead) to 1 (full health) health utility scale. We analyzed the correlation between panelists' mean importance ratings and utility ratings using the Pearson correlation coefficient (r). Analyses were completed using IBM SPSS Statistics 19 and Microsoft Excel.

3. Results

Health outcome descriptors developed for ASH VTE Guidelines

We developed 127 HODs for 104 outcomes for VTE guidelines (see Appendix B for the complete list). Figure 2 shows examples of the finalized HODs. Eighteen of the outcomes were described with multiple levels of severity, either with two levels (e.g. non-severe and severe) or three levels (e.g. mild, moderate, severe), while the remainder were described with a single HOD, i.e. one level of severity. The HODs for outcome importance rating and utility rating were drafted and finalized by 18 participating clinician experts or methodologists over a period of approximately 16 weeks. All HODs are available on a web-based registry (https://ms.gradepro.org/).

The anatomy of a health outcome descriptor

The HODs developed through our approach describe in a structured format the common symptoms experienced due to a health outcome, how long the health outcome lasts (time horizon), the tests and treatments that a person experiencing the outcome is expected to undergo and, finally, the consequences for a person experiencing the outcome including the prognosis, long term effects and side effects.

The structure and details of the HODs allow defining outcomes across a range of disease including describing different severities of an outcome that would result in different consequences for a person experiencing that outcome. What this means for guideline developers is that outcomes considered in decision-making that may typically be defined with a simple label (e.g. allergic reaction), but can have a broad range of consequences, can be more specifically defined with a HOD. Within the topic area of VTE, thrombosis outcomes such as DVT and pulmonary embolism (PE) would typically be considered as critical outcomes for decisions about prevention and treatment, as they were by our panels. However, not all occurrences of DVT and PE may have the same consequences for patients experiencing them, as well as the same implications for management and care options. For example, Figure 2 provides the HODs for the PE health outcome, with 3 levels of severity, showing for example that a PE of mild severity would commonly result in shortness of breath but could be treated at home, whereas a severe PE would require oxygen administration and hospitalization, frequently in a critical care unit.

Outcome prioritization and utility rating results using health outcome descriptors

The outcome importance rating and outcome utility rating surveys incorporating the HODs were completed by 111 of 118 (94%) and 79 of 118 panel members (67%), respectively. For each of the 18 outcomes with multiple severities, we found that panelists' mean ratings showed higher importance ratings (i.e. closer to being considered a critical outcome) and lower utility ratings (i.e. closer to a state of being dead) for outcomes of greater severity (see Table 4), suggesting that HODs could be interpreted by the panelists as expressing different consequences and levels of severity of an outcome. For example, importance ratings for PE had a mean (SD) of 6.7 (0.33), 7.9 (0.14), and 8.8 (0.12) for the mild, moderate, and severe HODs, respectively. The utility ratings for PE had a mean (SD) of 0.62 (0.16), 0.42 (0.15), 0.25 (0.14) for the mild, moderate, and severe HODs, respectively.

Comparing panels' mean importance and utility ratings for all HODs, we saw a strong correlation between the panels' importance and utility ratings (Pearson's r=-0.88, p<0.001), with outcomes rated higher on the 9-point importance scale receiving lower utility ratings (i.e. closer to 0 on the VAS) (see Figure 3). This demonstrates that using the two methods of measurement incorporating HODs, which were administered to the panels several months apart, we obtained similar findings on panelists' perspectives about how important specific outcomes are. As all panels rated the importance of the complete set of outcomes, we also saw good agreement between panels, with 82% of mean panel ratings falling in the same importance category according to the 9-point

scale. The panels' ratings for the full set of 127 HODs are reported in the companion publication.[6]

Use of HODs in the evidence-to-decision process

Use of the HODs we developed provided a reference point throughout the guideline development process when health outcomes were considered. HODs provided explicit definitions of outcomes when incorporated into importance and utility rating surveys, they could be incorporated into Summary-of-Findings tables presenting research evidence on effects of interventions on outcomes and in EtD frameworks when judging the size and balance of health benefits and harms (see Appendix C for examples of incorporating the HODs).

In the step of making judgements about evidence-to-decision criteria to arrive at a recommendation, panels were able to view the specific HODs to be considered for decision-making in the EtD framework (see Figure 4). This helped ensure that panelists were considering the same outcome (i.e. with the same consequences and severity) during their deliberations. To decide about the magnitude of desirable effects (or 'benefits') panels would need to judge how large those were considering the effect size for the HODs. When it came to deciding about the balance between desirable and undesirable effects (or 'benefits and harms') HODs were intended to ensure a common construct or understanding between panelists to allow for appropriate weighing. In the example in Figure 4, this guideline panel considered that the outcome of PE occurring in this population would typically be that of moderate severity, as opposed to mild or

severe severity, and prioritized this HOD. Considering an intervention that reduced PE but increased bleeding in patients, panelists could understand that they were weighing the trade-off of an intervention increasing major bleeding but reducing moderate PE, rather than mild or severe PE. A panel would presumably deem an intervention to be more favoured if it increased major bleeding but reduced moderate PE, and favoured less (or perhaps not favoured) if the PE it reduced was mild. Allowing a common mental construct of outcomes across panelists by utilizing a HOD aimed to avoid potential disagreement in judgements about magnitude or balance of effects simply due to panels having a different understanding or implicit definitions of the outcomes.

Face validity

The approach for developing HODs appeared to be effective as it allowed engaging panelists and methodologists in creating a large number of HODs following the instructional guide. It was possible to use the HODs developed to conduct our outcome importance and utility rating exercises, as well as to integrate the HODs into the evidence-to-decision process.

4. Discussion

Summary of Findings

We created a new approach to harmonize the consideration of health outcomes in guidelines that allows calibration of guideline developers and users to those outcomes. In this article we described the approach to developing HODs in detail and provide an instructional guide, the resulting features of HODs, and validity of applying

the approach in a guideline development project with ten panels. In our companion publication we describe how HODs can be used for prioritization in guideline development and report on the result of the panels' outcome prioritization and outcome utility rating using HODs.

We believe this work demonstrates feasibility of the approach by developing a set of 127 HODs, across different clinical albeit related topic areas. We incorporated the HODs in the development of ten guidelines. Panel members consistently rated the importance of HODs, including in many instances those outside of their specific area of interest, and later provided utility ratings that correlated with their outcome importance ratings. HODs provide a structured definition of the health outcomes considered by panels, making available a reference point for discussion of outcomes during panel deliberations. HODs also provide an approach of communicating judgements to the users of the guidelines regarding how the panel explicitly defined and considered outcomes. The development of HODs and demonstrating their use addresses a critical issue in guideline development as it connects guideline developers to those using the guidelines, thereby, enhancing transparency.[22,23]

Strengths and Limitations

In our study, we applied the approach for developing and using HODs in a real guideline development project to enable assessments of feasibility and applicability of the methods. This allows for transferability of the approach to other guidelines, and it has since been applied to other topic areas.[24] The work on developing the approach

involved design and careful planning that started before the guideline development, as it was identified as part of a research agenda for the guideline project. In this process, we also involved a methods advisory group with extensive expertise in guideline development to provide input on the approach. We were able to successfully engage experts across ten guideline panels in drafting HODs. This in turn generated a large number of HODs, including multiple HODs representing different outcome severities, and a large sample of panelists' ratings to inform our analyses.

In this study we did not evaluate differences in rating of importance and utility of outcomes with or without the HODs, or potential differences in decision-making by panels with and without HODs, and this is an area for future research to validate the approach. We also did not evaluate the impact of having two HOD templates, one for importance rating and the more detailed one for utility rating. In many instances the two types of HODs were similar and the need for both templates should be explored in future work. Another limitation is that while we involved the panel patient representatives in the outcome rating exercises using HODs, we did not directly involve patient representatives in the development of the HODs. As the objective is for HODs to be relevant to patients, approaches to involving patients in HOD development and obtaining feedback are of priority. While HODs were incorporated for outcome prioritization for all panels, the extent to which the guideline panels in our study applied the HODs in the evidence-to-decision process varied. For example, several panels deliberated about HODs to the extent of modelling assumptions about the distribution

of HODs (e.g. proportion that are severe and critical to patients versus mild and not important) to inform their decision-making, while others did not make such distinctions and dedicated less time to deliberation about HODs within the EtD frameworks. This merits future evaluation and qualitative study on what may impact panel buy-in and views on the importance of applying the approach. Lastly, in this study we describe the development and use HODs in only one broad clinical topic area (management of VTE), but we have also applied this approach in other guidelines that included broader involvement of patients.[22]

Implications for Clinicians and Guideline Developers

Use of HODs aims to support determining the importance of outcomes and deliberating about the balance of benefits and harms by providing a common reference point for panels' judgements about interventions and healthcare options. HODs can be developed through involvement of experts serving on guideline panels. HODs not only focus on what matters to those affected by an intervention, they also make the considered outcomes transparent and can serve to inform researchers to focus on what matters to decision makers and support communication with other decision makers, such as those conducting evidence syntheses and policymakers.[23] Given this focus, HODs could also be used as a tool for communication with patients by including them in decision aids. The intent is to enable linking of health outcomes through the use of HODs across different stakeholders within a given clinical or health policy topic area.

Implications for Research

The development and use of HODs in the production of practice guidelines offers an approach towards maintaining a common understanding of outcomes throughout the guideline process for the different stakeholders involved. Future research should evaluate potential differences in decision-making by panels about the importance of, as well as the recommendations that are formulated on the basis of the same research evidence, when using and not using HODs. This could be accomplished through randomized controlled trials comparing the two approaches. Further research should also investigate how users and the target audience of guidelines interpret and understand HODs and recommendations developed with the approach.

5. Conclusions

This study describes an approach for HOD development and provides instruction on how guideline developers may implement it. The approach may be useful to provide a reference point for a common understanding of outcomes considered in the development of a guideline. Future efforts in this area should focus on further validation of the development and use of HODs and their potential to improve transparency and harmony between producers of research, guideline developers, and users of guideline recommendations.

Funding/support: The development of the VTE guidelines was funded by the American Society of Hematology.

Role of the funding source: The funder supported the guideline development work, but did not have a role in designing methodology and reporting.

Financial Disclosures / Conflict of interest:

T.L.O has served as a consultant for Instrumentation Laboratory, and his institution has received research support on his behalf from Instrumentation Laboratory, Siemens, Stago, NIH, CDC, and PCORI.

All other authors declare no competing interests.

Authorship contribution statement: HJS was the principal investigator. HJS, RN, WW, were responsible for conceptualization of the research study. HJS and WW conceived the original HOD template used in the study. HJS, RN, WW developed the instructional guide for HOD development. RN, PD, AI, RAM, IN, BR, VM, TLO, SMB, AC, WL, PM, RK, DMW, SRK, CM, SMR, NAZ, and HJS drafted HODs, WW was responsible for carrying out the data analysis. WW drafted the manuscript with critical revisions and writing contributions from HJS. All of the authors revised the manuscript critically for important intellectual content and approved the final version submitted for publication.

Acknowledgements: The authors thank the ASH VTE guidelines panel members for participating in the outcome rating exercises as part of the guideline development project.

References

1. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13(1):132.
2. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of clinical epidemiology.* 2011;64(4):395-400.
3. Atkins D, Perez-Padilla R, Macnee W, et al. Priority setting in guideline development: article 2 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proceedings of the American Thoracic Society.* 2012;9(5):225-228.
4. 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.
5. Alonso-Coello P, Schunemann 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.
6. Wiercioch W., Zhang Y., Nieuwlaat R., et al. New methods facilitated the process of prioritizing questions and health outcomes in guideline development *Journal of clinical epidemiology.* 2020;Submitted.
7. Anderson DR, Morgano GP, Bennett C, et al. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. *Blood Advances.* 2019;3(23):3898-3944.
8. Bates SM, Rajasekhar A, Middeldorp S, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: venous thromboembolism in the context of pregnancy. *Blood Advances.* 2018;2(22):3317-3359.
9. Cuker A, Arepally GM, Chong BH, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. *Blood Advances.* 2018;2(22):3360-3392.
10. Lim W, Le Gal G, Bates SM, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. *Blood Advances.* 2018;2(22):3226-3256.
11. Monagle P, Cuello CA, Augustine C, et al. American Society of Hematology 2018 Guidelines for management of venous thromboembolism: treatment of pediatric venous thromboembolism. *Blood Advances.* 2018;2(22):3292-3316.
12. Schünemann HJ, Cushman M, Burnett AE, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. *Blood Advances.* 2018;2(22):3198-3225.
13. Witt DM, Nieuwlaat R, Clark NP, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: optimal management of anticoagulation therapy. *Blood Advances.* 2018;2(22):3257-3291.
14. Schunemann HJ, Wiercioch W, Etxeandia I, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ :*

Canadian Medical Association journal = journal de l'Association medicale canadienne. 2014;186(3):E123-142.

15. Wiercioch W, Nieuwlaat R, Akl EA, et al. Methodology for the American Society of Hematology VTE guidelines: current best practice, innovations, and experiences. *Blood Advances.* 2020;4(10):2351-2365.
16. Schunemann HJ, Armstrong D, Degl'innocenti A, et al. A randomized multicenter trial to evaluate simple utility elicitation techniques in patients with gastroesophageal reflux disease. *Med Care.* 2004;42(11):1132-1142.
17. Schunemann HJ, Goldstein R, Mador MJ, et al. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. *Eur Respir J.* 2005;25(1):31-40.
18. Schunemann HJ, Stahl E, Austin P, Akl E, Armstrong D, Guyatt GH. A comparison of narrative and table formats for presenting hypothetical health states to patients with gastrointestinal or pulmonary disease. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2004;24(1):53-60.
19. Schünemann HJ, Griffith L, Stubbing D, Goldstein R, Guyatt GH. A clinical trial to evaluate the measurement properties of 2 direct preference instruments administered with and without hypothetical marker states. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2003;23(2):140-149.
20. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol.* 2013;66(2):158-172. doi: 110.1016/j.jclinepi.2012.1001.1012. Epub 2012 May 1018.
21. 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: 110.1016/j.jclinepi.2012.1008.1001. Epub 2012 Oct 1030.
22. Baldeh T, Saz-Parkinson Z, Muti P, et al. Development and use of health outcome descriptors: a guideline development case study. *Health Qual Life Outcomes.* 2020;18(1):167-167.
23. Schünemann HJ, Tunis S, Nieuwlaat R, Wiercioch W, Baldeh T. The SOLAR (Standardized Outcomes Linking Across StakeholdeRs) system and hub and spokes model for direct core outcome measures in health care and its relation to GRADE. *Journal of clinical epidemiology.* 2020.
24. Sekeres MA, Guyatt G, Abel G, et al. American Society of Hematology 2020 guidelines for treating newly diagnosed acute myeloid leukemia in older adults. *Blood Advances.* 2020;4(15):3528-3549.

Chapter 2: Tables

Table 1: Rating of HODs of outcomes with multiple severities

Outcome	Health Outcome Descriptor	Importance Rating – all panels (n=111)	Importance Rating – VTE in Medical Patients (n=12)	Importance Rating – VTE in Surgical Patients (n=13)	Importance Rating – VTE Treatment (n=12)	Importance Rating – Optimal Management (n=10)	Importance Rating – HIT (n=11)	Importance Rating – Thrombophilia (n=7)	Importance Rating – VTE in Cancer (n=15)	Importance Rating – Pediatric VTE (n=13)	Importance Rating – VTE in Pregnancy (n=10)	Importance Rating – Diagnosis of VTE (n=8)	Utility Rating – all panels (n=70)*
		Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD)
Acute Pain	Acute Pain - Mild	3.7 (0.49) Important	3.3 (1.48) Not Important	3.4 (1.87) Not Important	3.4 (1.32) Not Important	4.5 (1.69) Important	4.2 (1.34) Important	3.3 (1.67) Not Important	4.6 (1.2) Important	3.6 (1.27) Important	3.4 (1.5) Not Important	3.9 (1.62) Important	0.78 (0.14)
	Acute Pain - Moderate	5.5 (0.46) Important	5.1 (1.44) Important	5.4 (1.92) Important	5.5 (1.44) Important	6.4 (1.11) Important	6 (1.04) Important	5.1 (1.73) Important	6 (1.1) Important	4.8 (0.86) Important	5.3 (1.42) Important	5.4 (1.41) Important	0.55 (0.18)
	Acute Pain - Severe	6.8 (0.63) Critical	6.7 (1.7) Critical	6.5 (2.23) Important	7 (1.41) Critical	7.7 (1.19) Critical	7.4 (1.07) Critical	6 (1.77) Important	7.5 (1.09) Critical	5.6 (1.33) Important	6.6 (1.85) Critical	7.1 (1.17) Critical	0.35 (0.19)
Allergic Reaction to Contrast Dye	Allergic Reaction to Contrast Dye - Mild	3.2 (0.44) Not Important	3 (1.6) Not Important	2.8 (1.4) Not Important	3 (1.68) Not Important	4.4 (1.8) Important	3.3 (1.54) Not Important	2.9 (1.88) Not Important	3.5 (1.54) Not Important	3.1 (1.11) Not Important	3 (1.18) Not Important	3 (1.32) Not Important	0.84 (0.15)
	Allergic Reaction to Contrast Dye - Moderate	4.6 (0.54) Important	4.4 (2.14) Important	4.4 (1.61) Important	4.1 (1.93) Important	5.8 (1.66) Important	4.9 (1.68) Important	3.7 (1.98) Important	4.8 (1.38) Important	4.8 (1.42) Important	4.6 (1.91) Important	4.3 (1.2) Important	0.71 (0.18)
	Allergic Reaction to Contrast Dye - Severe	5.9 (0.35) Important	5.8 (2.17) Important	5.8 (2.08) Important	5.7 (2.39) Important	6.5 (1.57) Critical	6.1 (2.27) Important	5.1 (2.17) Important	6.2 (1.56) Important	5.9 (1.14) Important	6.2 (2.36) Important	5.8 (1.64) Important	0.53 (0.24)
Cerebral Venous Thrombosis	Cerebral Venous Thrombosis – Mild	7.1 (0.47) Critical	7.1 (0.64) Critical	6.8 (1.54) Critical	5.9 (1.66) Important	7.5 (1.02) Critical	7.2 (1.34) Critical	7.3 (1.48) Critical	6.9 (1.39) Critical	7.7 (0.91) Critical	7.4 (1.28) Critical	6.9 (1.54) Critical	0.47 (0.18)
	Cerebral Venous Thrombosis – Severe	8.3 (0.41) Critical	8.6 (0.48) Critical	8.1 (1.14) Critical	7.5 (1.67) Critical	8.7 (0.46) Critical	7.8 (1.4) Critical	8.4 (0.73) Critical	8.1 (1.59) Critical	8.8 (0.36) Critical	8.6 (0.66) Critical	8.3 (0.66) Critical	0.24 (0.15)
Chronic Pain	Chronic Pain - Mild	4.5 (0.54) Important	3.9 (1.55) Important	4.1 (2.27) Important	4.5 (1.12) Important	5 (1.84) Important	4.6 (1.07) Important	5.7 (1.75) Important	4.6 (1.54) Important	4.8 (1.56) Important	4 (1.61) Important	4 (1.58) Important	0.68 (0.16)
	Chronic Pain - Moderate	6.2 (0.5) Important	6.1 (1.75) Important	5.3 (2.3) Important	6.3 (1.01) Important	7.1 (1.37) Critical	6.5 (1.08) Important	6.9 (2.17) Critical	6.3 (1.44) Important	6.2 (1.31) Important	6.2 (1.89) Important	5.6 (1.22) Important	0.45 (0.19)
Cognitive Impairment	Cognitive Impairment - Mild	5.8 (0.56) Important	5.8 (1.86) Important	5.3 (2.26) Important	4.9 (1.66) Important	5.9 (1.45) Important	6.8 (1.34) Critical	5.6 (1.84) Important	5.5 (1.41) Important	6 (2.39) Important	6.6 (2.24) Critical	5.4 (1.8) Important	0.47 (0.21)
	Cognitive Impairment - Severe	7.3 (0.54) Critical	7.5 (1.85) Critical	6.5 (2.39) Important	6.8 (1.36) Critical	7.6 (1.8) Critical	8.3 (1.35) Critical	7.6 (1.18) Critical	6.7 (1.81) Critical	7.6 (1.9) Critical	7.8 (2.32) Critical	7.3 (1.71) Critical	0.22 (0.18)
DVT in Forearm	Deep Venous Thrombosis (DVT) in the Forearm – Mild	4.7 (0.75) Important	4.5 (1.04) Important	3.6 (1.6) Important	3.6 (2.18) Important	6.2 (1.4) Important	4.5 (1.72) Important	4.9 (2.1) Important	4.9 (2.05) Important	5.5 (1.65) Important	4.9 (2.47) Important	4.5 (2.24) Important	0.78 (0.15)
	Deep Venous Thrombosis (DVT) in the Forearm – Moderate	5.6 (0.66) Important	5.5 (0.87) Important	5.1 (1.77) Important	4.8 (1.92) Important	7.1 (1.66) Critical	5.4 (1.07) Important	5.6 (2.06) Important	5.7 (1.81) Important	6.4 (1.39) Important	5.9 (2.12) Important	5 (2.18) Important	0.68 (0.18)

	Deep Venous Thrombosis (DVT) in the Forearm – Severe	6.4 (0.59) Important	6.4 (1.11) Important	5.8 (1.87) Important	5.7 (1.49) Important	7.5 (1.57) Critical	6.3 (0.86) Important	6.3 (1.98) Important	6.6 (1.74) Critical	7.3 (1.26) Critical	6.8 (1.54) Critical	5.8 (1.98) Important	0.56 (0.2)
DVT in Lower Leg (Distal)	Deep Venous Thrombosis (DVT) in the Lower Leg – Mild	5.3 (0.71) Important	4.9 (1.19) Important	4.7 (1.49) Important	3.9 (2.14) Important	6.6 (1.5) Critical	5.2 (1.64) Important	5.4 (1.59) Important	5.7 (1.44) Important	6 (1.75) Important	5.7 (1.95) Important	5.1 (2.03) Important	0.77 (0.15)
	Deep Venous Thrombosis (DVT) in the Lower Leg – Moderate	6.4 (0.56) Important	6.1 (0.95) Important	6.1 (1.44) Important	5.3 (1.83) Important	7.5 (1.2) Critical	6.3 (1.54) Important	6.4 (1.18) Important	6.7 (1) Critical	6.7 (1.43) Critical	6.8 (1.47) Critical	6.1 (1.9) Important	0.64 (0.16)
	Deep Venous Thrombosis (DVT) in the Lower Leg – Severe	7.2 (0.5) Critical	7 (0.71) Critical	7.2 (1.17) Critical	6.2 (1.77) Important	8.1 (1.14) Critical	7.1 (1.16) Critical	7.4 (1.18) Critical	7.5 (0.91) Critical	7.5 (1.04) Critical	7.6 (1.07) Critical	6.7 (1.48) Critical	0.52 (0.17)
DVT in Upper Arm	Deep Venous Thrombosis (DVT) in the Upper Arm – Mild	5.6 (0.61) Important	5.4 (1.19) Important	5 (1.36) Important	4.3 (2.05) Important	6.6 (1.43) Critical	5.4 (1.67) Important	5.7 (1.75) Important	5.7 (1.66) Important	6.3 (1.26) Important	5.9 (2.02) Important	5.5 (2) Important	0.75 (0.15)
	Deep Venous Thrombosis (DVT) in the Upper Arm – Moderate	6.6 (0.42) Critical	6.6 (0.86) Critical	6.4 (0.84) Important	5.8 (1.48) Important	7.3 (1.27) Critical	6.5 (1.23) Critical	6.4 (1.76) Important	6.7 (1.19) Critical	7.2 (0.86) Critical	6.8 (1.6) Critical	6.3 (1.56) Important	0.61 (0.16)
	Deep Venous Thrombosis (DVT) in the Upper Arm – Severe	7.5 (0.35) Critical	7.5 (0.65) Critical	7.5 (0.84) Critical	6.8 (0.9) Critical	7.9 (1.14) Critical	7.5 (0.78) Critical	7.4 (1.59) Critical	7.3 (1.19) Critical	8.2 (0.58) Critical	7.6 (1.02) Critical	7.4 (1.22) Critical	0.48 (0.17)
DVT in Upper Leg (Proximal)	Deep Venous Thrombosis (DVT) in the Upper Leg – Mild	6.1 (0.5) Important	5.9 (1.38) Important	5.7 (1.73) Important	5.2 (1.75) Important	7 (1.1) Critical	6.2 (1.8) Important	5.7 (1.67) Important	6.6 (0.88) Critical	6.3 (1.14) Important	6.6 (1.56) Critical	6.1 (1.76) Important	0.71 (0.17)
	Deep Venous Thrombosis (DVT) in the Upper Leg – Moderate	7.1 (0.41) Critical	7 (1) Critical	7.2 (1.41) Critical	6.5 (1.19) Critical	7.8 (0.87) Critical	6.8 (1.34) Critical	6.6 (1.59) Critical	7.5 (0.62) Critical	7.5 (1.01) Critical	7.6 (1.28) Critical	7.1 (1.36) Critical	0.58 (0.14)
	Deep Venous Thrombosis (DVT) in the Upper Leg – Severe	8 (0.27) Critical	8.3 (0.43) Critical	8 (0.96) Critical	7.8 (0.83) Critical	8.4 (0.66) Critical	7.7 (1.21) Critical	7.9 (0.83) Critical	8.2 (0.65) Critical	8.4 (0.74) Critical	8.2 (0.75) Critical	7.6 (1.22) Critical	0.43 (0.16)
Gastrointestinal Bleeding	Gastrointestinal Bleeding - Major	7.4 (0.42) Critical	7.4 (1.38) Critical	7.3 (1.1) Critical	7.7 (1.18) Critical	8.1 (1.04) Critical	7.4 (1.15) Critical	7 (1.07) Critical	7.9 (0.68) Critical	6.8 (0.97) Critical	7.8 (0.87) Critical	7 (0.71) Critical	0.44 (0.19)
	Gastrointestinal Bleeding - Minor	5.1 (0.53) Important	5.3 (1.42) Important	4.9 (1.97) Important	4.5 (1.62) Important	6.4 (1.43) Important	5.3 (0.86) Important	5 (1.6) Important	5.1 (1.75) Important	4.5 (1.28) Important	5.5 (1.63) Important	5 (1.32) Important	0.71 (0.16)
Infant Bleeding	Infant Bleeding - Mild	4.9 (0.61) Important	4.5 (1.5) Important	4.6 (1.96) Important	5.2 (1.85) Important	6.3 (1.94) Important	4.8 (1.2) Important	4.4 (2.15) Important	4.5 (1.56) Important	4.2 (1.35) Important	5.4 (1.36) Important	5.3 (1.71) Important	0.67 (0.21)
	Infant Bleeding - Severe	8.1 (0.75) Critical	8.2 (0.69) Critical	8.4 (0.68) Critical	8.3 (0.75) Critical	8.7 (0.67) Critical	8.1 (0.6) Critical	6 (2.9) Important	7.9 (1.98) Critical	8.1 (0.92) Critical	8.9 (0.3) Critical	8.3 (0.83) Critical	0.26 (0.19)
Ischemic Stroke	Ischemic Stroke - Mild	7 (0.61) Critical	6.3 (1.89) Important	6.3 (1.35) Important	6.4 (1.5) Important	7.3 (1.19) Critical	7.6 (0.98) Critical	8 (0.93) Critical	6.5 (1.5) Critical	6.8 (0.97) Critical	7.8 (0.75) Critical	7 (1.41) Critical	0.39 (0.19)
	Ischemic Stroke - Severe	8.3 (0.39) Critical	7.8 (2.07) Critical	8.2 (1.03) Critical	7.6 (1.93) Critical	8.6 (0.66) Critical	8.6 (0.48) Critical	8.9 (0.35) Critical	8.1 (1.31) Critical	8.2 (1.12) Critical	8.8 (0.4) Critical	8.4 (0.7) Critical	0.14 (0.1)
Bleeding	Major Bleeding	8.4 (0.3) Critical	8.7 (0.47) Critical	8.2 (0.58) Critical	8.3 (0.75) Critical	8.7 (0.64) Critical	8.5 (0.5) Critical	7.7 (0.7) Critical	8.5 (0.62) Critical	8.2 (0.8) Critical	8.8 (0.4) Critical	8.3 (0.97) Critical	0.33 (0.23)
	Minor Bleeding	4.5 (0.68) Important	4.2 (1.57) Important	4.4 (1.44) Important	4.4 (1.61) Important	5.7 (1.42) Important	4.8 (1.75) Important	3.1 (1.88) Not Important	5.3 (1.53) Important	3.8 (1.31) Important	4.5 (1.63) Important	4.6 (1.22) Important	0.81 (0.15)

Anxiety	Mild Anxiety	3.5 (0.58) Important	3.5 (1.44) Important	2.7 (1.71) Not Important	3.4 (1.85) Not Important	4.5 (1.96) Important	2.7 (0.96) Not Important	4.3 (2.05) Important	4.1 (1.65) Important	3.7 (1.2) Important	3.1 (1.58) Not Important	3.4 (1.11) Not Important	0.84 (0.11)
	Moderate to Severe Anxiety	5 (0.61) Important	5.3 (1.93) Important	3.7 (1.71) Important	4.5 (1.88) Important	5.7 (1.55) Important	4.4 (1.37) Important	5.9 (1.73) Important	5.1 (1.86) Important	5.2 (1.17) Important	5.1 (1.51) Important	4.9 (1.45) Important	0.65 (0.18)
Neonatal Bleeding	Neonatal Bleeding - Mild	4.8 (0.72) Important	4.8 (1.88) Important	4.1 (1.58) Important	5.5 (1.97) Important	6.3 (1.94) Important	4.1 (1.36) Important	4.2 (2.4) Important	4.6 (1.49) Important	4 (1.52) Important	5 (1.1) Important	5.4 (1.49) Important	0.65 (0.23)
	Neonatal Bleeding - Severe	7.9 (0.43) Critical	8.3 (0.72) Critical	7.8 (0.87) Critical	8.2 (0.94) Critical	8.6 (0.68) Critical	7.8 (1.09) Critical	7 (2.1) Critical	7.7 (2.05) Critical	7.5 (1.28) Critical	8.3 (0.64) Critical	8 (0.87) Critical	0.3 (0.21)
Placental Abruption	Placental Abruption – Non-Severe	5.1 (0.46) Important	5.6 (0.95) Important	4.3 (1.66) Important	5.2 (1.64) Important	5.9 (2.18) Important	4.6 (1.49) Important	5 (2) Important	5.3 (1.49) Important	4.7 (1.81) Important	5.4 (2.5) Important	5.4 (1.32) Important	0.69 (0.19)
	Placental Abruption – Severe	6.8 (0.46) Critical	6.8 (1.34) Critical	6.3 (2.26) Important	7.1 (1.38) Critical	6.8 (1.81) Critical	6.9 (1.9) Critical	6.4 (1.99) Important	7.2 (1.86) Critical	5.8 (2.32) Important	7 (2.28) Critical	7.5 (1) Critical	0.48 (0.24)
Pulmonary Embolism	Pulmonary Embolism - Mild	6.7 (0.33) Critical	6.5 (1.32) Critical	6.2 (1.67) Important	6.3 (1.88) Important	7.2 (1.03) Critical	6.5 (1.62) Critical	6.6 (1.92) Critical	7.1 (0.77) Critical	6.7 (1.2) Critical	7.1 (1.37) Critical	6.6 (1.8) Critical	0.62 (0.16)
	Pulmonary Embolism - Moderate	7.9 (0.14) Critical	7.9 (0.64) Critical	7.8 (1.12) Critical	7.9 (0.86) Critical	8.2 (0.6) Critical	7.9 (0.9) Critical	7.7 (1.28) Critical	7.9 (0.72) Critical	7.9 (0.62) Critical	8.1 (0.94) Critical	8 (0.71) Critical	0.42 (0.15)
	Pulmonary Embolism - Severe	8.8 (0.12) Critical	8.8 (0.43) Critical	8.8 (0.58) Critical	8.9 (0.28) Critical	8.8 (0.4) Critical	8.8 (0.4) Critical	8.4 (0.73) Critical	8.8 (0.4) Critical	8.8 (0.36) Critical	8.7 (0.46) Critical	8.8 (0.43) Critical	0.25 (0.14)
Acne**	Acne - Mild	-	-	-	-	-	-	-	-	-	-	-	0.89 (0.13)
	Acne - Severe	-	-	-	-	-	-	-	-	-	-	-	0.75 (0.19)

* 9 of 79 panelists completing the utility rating survey used the VAS in reverse, and were excluded from analysis

** The acne outcome was suggested as missing in the outcome rating importance survey and it was, therefore, included only in the subsequent step of utility rating.

Chapter 2: Figures

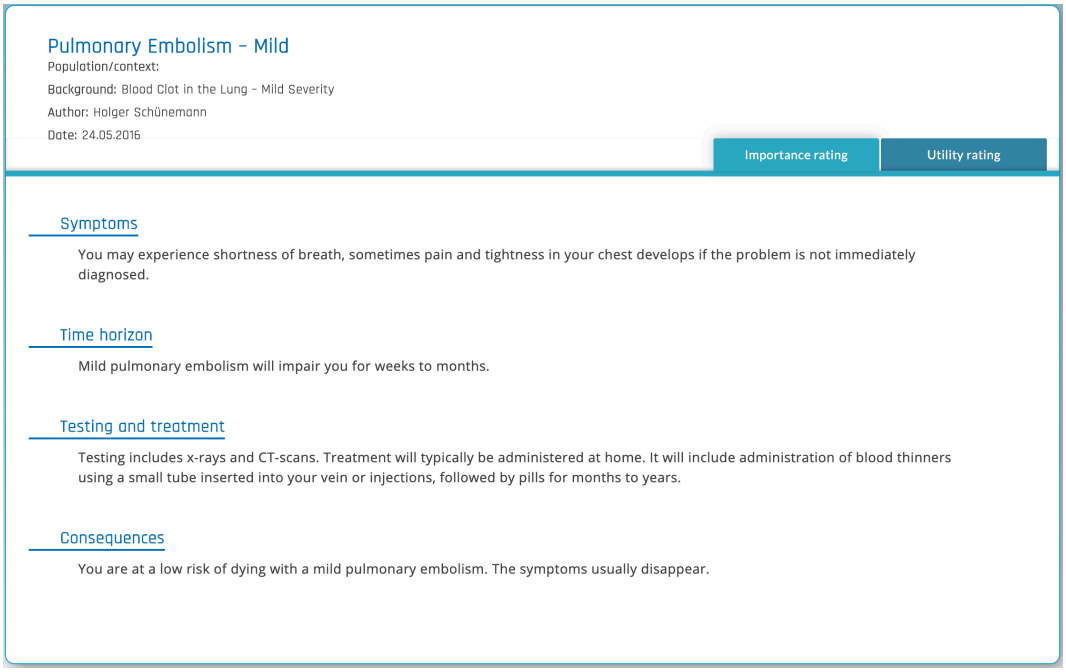
Figure 1. Template for a health outcome descriptor

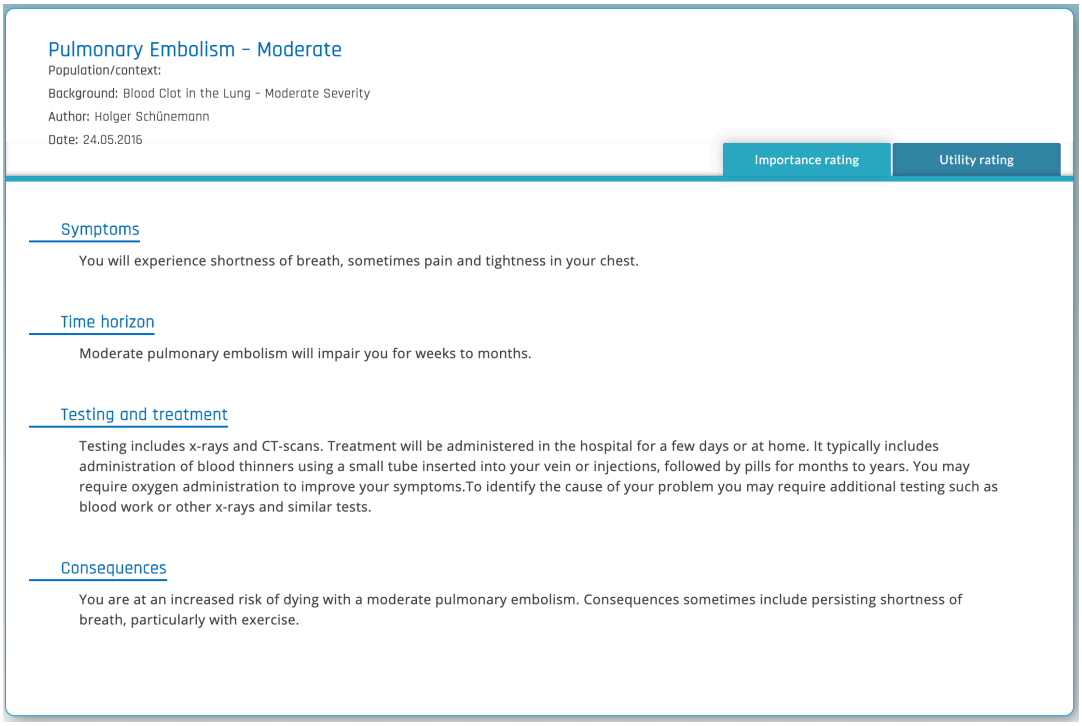
[Health Outcome Descriptor Title]

[Health outcome descriptor lay title]

- **Symptoms:** Common symptoms due to the health state. Note that grade of severity can be labeled into mild, moderate or severe and will be used as a descriptor of the health outcome descriptor, not as part of the symptoms. ["You experience xx…"]
- **Time Horizon:** Within which timeframe does the health state occur. [" xx will persist for [hours/days/months] and will xx improve…]
- **Testing and Treatment:** Which tests and treatments are commonly applied for this health state. ["Treatment may be administered xx… Treatment typically includes xx…"]
- **Consequences:** including prognosis and side effects. ["Include xx…"]

Figure 2. Example of health outcome descriptors





Pulmonary Embolism – Moderate

Population/context:

Background: Blood Clot in the Lung – Moderate Severity

Author: Holger Schünemann

Date: 24.05.2016

Importance rating | Utility rating

Symptoms

You will experience shortness of breath, sometimes pain and tightness in your chest.

Time horizon

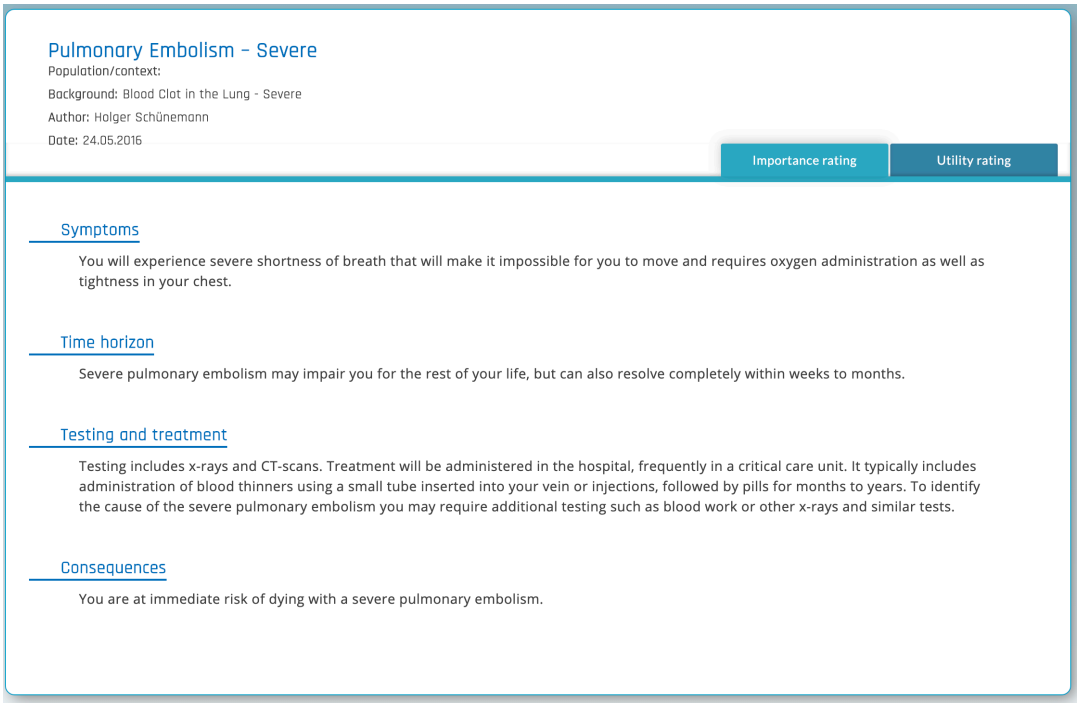
Moderate pulmonary embolism will impair you for weeks to months.

Testing and treatment

Testing includes x-rays and CT-scans. Treatment will be administered in the hospital for a few days or at home. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. You may require oxygen administration to improve your symptoms.To identify the cause of your problem you may require additional testing such as blood work or other x-rays and similar tests.

Consequences

You are at an increased risk of dying with a moderate pulmonary embolism. Consequences sometimes include persisting shortness of breath, particularly with exercise.



Pulmonary Embolism – Severe

Population/context:

Background: Blood Clot in the Lung - Severe

Author: Holger Schünemann

Date: 24.05.2016

Importance rating | Utility rating

Symptoms

You will experience severe shortness of breath that will make it impossible for you to move and requires oxygen administration as well as tightness in your chest.

Time horizon

Severe pulmonary embolism may impair you for the rest of your life, but can also resolve completely within weeks to months.

Testing and treatment

Testing includes x-rays and CT-scans. Treatment will be administered in the hospital, frequently in a critical care unit. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. To identify the cause of the severe pulmonary embolism you may require additional testing such as blood work or other x-rays and similar tests.

Consequences

You are at immediate risk of dying with a severe pulmonary embolism.

Figure 3: Relationship between panels' outcome importance ratings on 9-point scale (not important to critical) vs. outcome utility ratings on a 0 (dead) to 1 (full health) scale.

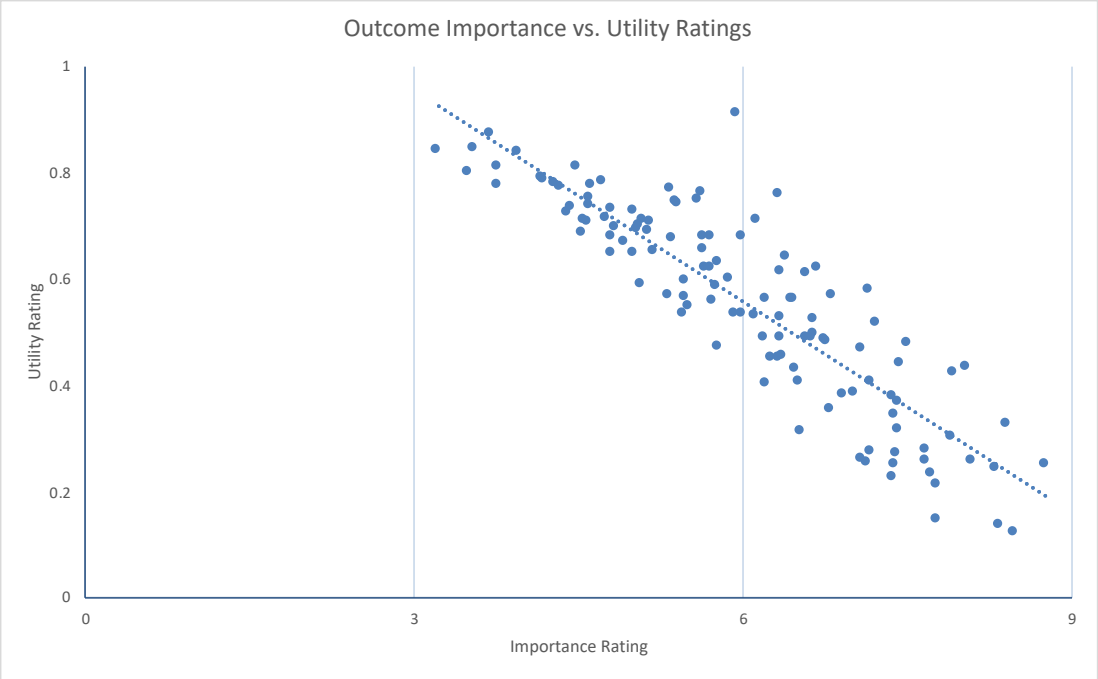
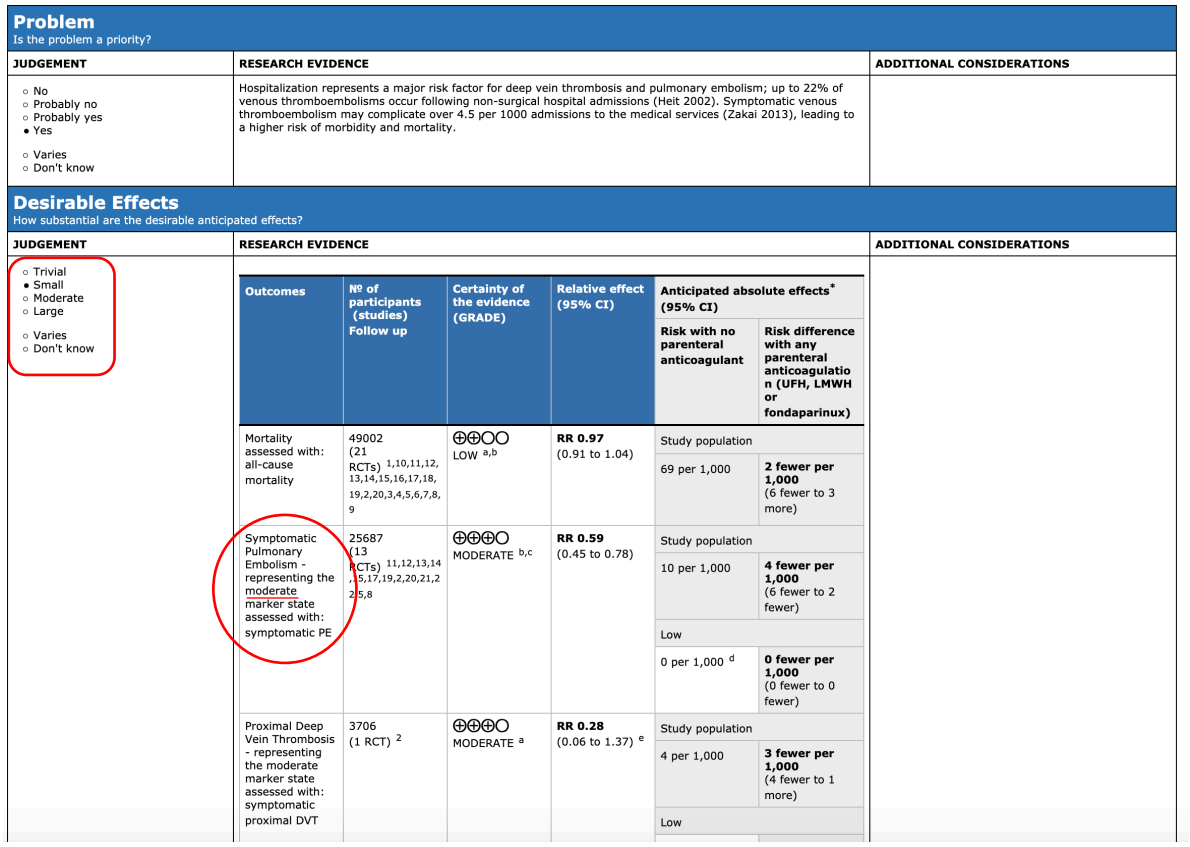


Figure 4: Evidence-to-decision frameworks



Problem
Is the problem a priority?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Hospitalization represents a major risk factor for deep vein thrombosis and pulmonary embolism; up to 22% of venous thromboembolisms occur following non-surgical hospital admissions (Heit 2002). Symptomatic venous thromboembolism may complicate over 4.5 per 1000 admissions to the medical services (Zakai 2013), leading to a higher risk of morbidity and mortality.	

Desirable Effects
How substantial are the desirable anticipated effects?

JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
○ Trivial ● Small ○ Moderate ○ Large ○ Varies ○ Don't know		

Outcomes	№ of participants (studies) Follow up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with no parenteral anticoagulant	Risk difference with any parenteral anticoagulation (UFH, LMWH or fondaparinux)
Mortality assessed with: all-cause mortality	49002 (21 RCTs) 1,10,11,12,13,14,15,16,17,18,19,2,20,3,4,5,6,7,8,9	⨁⨁◯◯ LOW a,b	**RR 0.97** (0.91 to 1.04)	Study population	
				69 per 1,000	**2 fewer per 1,000** (6 fewer to 3 more)
Symptomatic Pulmonary Embolism - representing the moderate marker state assessed with: symptomatic PE	25687 (13 RCTs) 11,12,13,14,15,17,19,2,20,21,22,5,8	⨁⨁⨁◯ MODERATE b,c	**RR 0.59** (0.45 to 0.78)	Study population	
				10 per 1,000	**4 fewer per 1,000** (6 fewer to 2 fewer)
				Low	
				0 per 1,000 d	**0 fewer per 1,000** (0 fewer to 0 fewer)
Proximal Deep Vein Thrombosis - representing the moderate marker state assessed with: symptomatic proximal DVT	3706 (1 RCT) 2	⨁⨁⨁◯ MODERATE a	**RR 0.28** (0.06 to 1.37) e	Study population	
				4 per 1,000	**3 fewer per 1,000** (4 fewer to 1 more)
				Low	

Figure 4 Caption: In the step of making judgements about evidence-to-decision criteria to arrive at a recommendation, panels were able to view the specific HODs to be considered for decision-making.

Chapter 2: Appendix A - Health outcome descriptor development instructional guide for panelists

Health Outcome Descriptor Development Guide

Purpose of this document
To provide assistance to chairs, co-chairs and panelists of ASH VTE guidelines in developing health outcome descriptors (HODs) (also referred to as marker states) for Outcome Importance and Outcome Utility rating with guideline panel members.

Definitions

Health outcome descriptor
A narrative or point-by-point description of a health state (or outcome) that facilitates understanding its key attributes.

Outcome importance rating
An initial rating of the relative importance or weight raters place on a health state (or outcome) on a 1 to 9 scale (1-3 = low importance for decision making, 4-6 = important, but not critical for decision making, 7-9 = critical for decision making). It serves to determine which outcomes to include in GRADE Evidence Profiles for panels (those rated important or critical for decision making), which will be the outcomes considered when the panel formulates a recommendation.

Outcome utility rating
A detailed economic theory-derived relative importance rating on a 0 to 100 scale, that can be used for decision and economic modeling.

Anticipated products of this work

- HODs will be developed for all patient-important outcomes initially suggested by ASH panels under the guidance of MacGRADE and according to this guide
- The complete list of outcomes across panels will be compiled by the MacGRADE group
- To develop a core outcome set for VTE guidelines, based on the most critical outcomes
- To establish visual analogue scales (VAS) as preferred measurement tool for outcome utility ratings in the guideline field/community

Methods
Level of comprehension

- HODs should be understood by non-specialized health workers and researchers, and by patients in general

Information sources for HOD content for this project

- Existing HODs
- Items from Health-Related Quality of Life measurement tools
- (Specialist) clinical observations

Outcome importance rating HODs

- HODs should consist of 4 bullet points, with one bullet point for each domain: symptoms, time horizon, testing and treatment, consequences. We will label the bullet points according to these 4 domains
- These should be the most essential characteristics that describe the 'average' (or usual) representation of the health state and cover the relevant domains.
- The four domains should describe the following:
- ✓ Symptoms: common symptoms due to the health state. Note that grade of severity can be labeled into mild, moderate or severe and will be used as a descriptor of the health marker state, not as part of the symptoms
- ✓ Time horizon: within which timeframe does the health state occur
- ✓ Testing & Treatment: which tests and treatments are commonly applied for this health state
- ✓ Consequences: including prognosis and side effects
- If a detailed HOD is already available, we can complete that HOD and condense to the 4 most important items based on the domains
- The health state description does not need to cover the full possible range of disease
- Write at a Grade 8 readability level, and use active and personal language ("you")
- See *Appendix I* for the template to develop a HOD for outcome importance rating
- See *Appendix II* for an example HOD, and a description of how to rate the outcome importance.
- A description of the health state (explanation of what it is) is not part of the bullet points but should be provided separately.

Outcome utility rating HODs

- These HODs **expand** on the Outcome importance rating HOD
- Items in the following domains: symptoms, time horizon, testing and treatment, consequences
- Includes up to 10 bullet points in total, with 2 to 4 per domain. Label the bullet points according to the 4 domains (symptoms, time horizon, testing and treatment, consequences)

- Should include a labeling of the severity of the HOD as part of the outcome definition (e.g. severe DVT, severe bleed, minor bleed) which is not part of the bullet points
- Should include sufficient details for patients to determine utility compared with perfect health and death. Write at a Grade 8 readability level, and use active and personal language ("you")
- See *Appendix I* for the template to develop a HOD for outcome utility rating
- See *Appendix III* for an example HOD, and a description of how to rate the outcome utility using a VAS scale
- A description of the health state (explanation of what it is) is not part of the bullet points but should be provided separately

See following pages for examples.

Appendix I: Template for HOD Development for Outcome Importance & Utility Rating

Title – importance rating

- **Symptoms:** You experience xxx
- **Time Horizon:** xxx will persist for [hours/days/months] and will xxx improve
- **Testing and Treatment:** Treatment may be administered xxx. Treatment typically includes xxx
- **Consequences:** Include xxx.

Title – utility rating

Symptoms:

- You experience xxx;
- xxx

Time Horizon:

- xxx will persist for [hours/days/months] and will xxx improve

Testing and Treatment:

- You will require testing with xxx
- You will receive xxx
- Your treatment will typically include administration xxx
- You may have to take xxx

Consequences:

- Consequences may include xxx
- You may be xxx
- …

Appendix II: Example of HOD and Rating Approach for Outcome Importance Rating Severe Deep Venous Thrombosis (DVT) in the Upper Leg – importance rating

- **Symptoms:** You experience severe swelling, pain, warmth, heaviness or redness in your entire leg.
- **Time Horizon:** Severe DVT will persist for months and will slowly improve.
- **Testing and Treatment:** Treatment may be administered in the hospital or at home. Treatment typically includes administration of blood thinners using an intravenous line, injections or pills. Long-lasting treatment with blood thinners is often required.
- **Consequences:** Consequences often include long-lasting pain and swelling in the leg. Sometimes, it may also include a blood clot travelling to the lungs (a pulmonary embolism) and death.

<u>Rating approach</u>

Many outcomes may be considered important by different patients or health care providers. However, not all outcomes are critical or important for making clinical decisions.

Using the following scale, please rate how important the outcome of 'Severe DVT in the leg' is for making a clinical decision about the optimal management strategy:

rating scale:								
1	2	3	4	5	6	7	8	9
of **least** importance								of **most** importance
of limited importance for making a decision (not included in evidence profile)			important, but not critical for making a decision (included in evidence profile)			Critical for making a decision (included in evidence profile)		

Appendix III: Example of HOD and Rating Approach for Outcome Utility Rating

Severe Deep Venous Thrombosis (DVT) in the Upper Leg – outcome utility rating
Symptoms:

- You experience severe swelling, pain, warmth, heaviness or redness in the entire leg;
- Walking is very uncomfortable and/or painful for you.

Time Horizon:

- The severe DVT in your leg will persist for months and slowly improve.

Testing and Treatment:

- You will require blood tests or radiological tests that are not painful.
- You will receive immediate initial treatment in hospital.
- Your treatment will typically include administration of blood thinners using an intravenous line initially, followed by daily injections or pills.
- You may have to take blood thinner pills for the rest of your life.

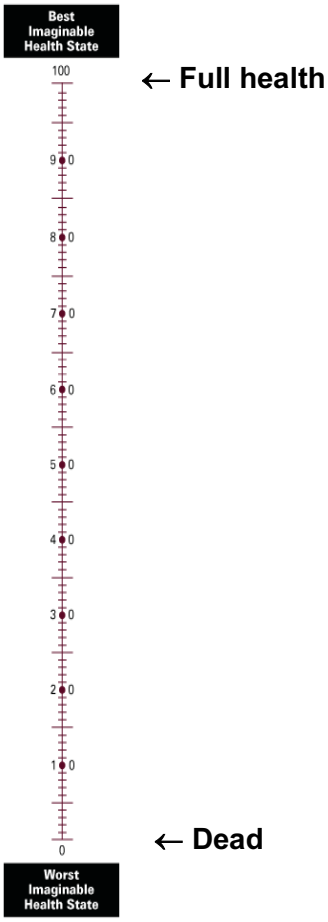
Consequences:

- Consequences often include chronic pain and swelling in your leg, and sometimes a severe complication such as a blood clot displacing to your lungs (pulmonary embolism) which can lead to death.
- You may be worried about side effects, including bleeding as a result of taking blood thinners.
- To reduce swelling you wear compression stockings most days. These stockings are difficult to put on and somewhat uncomfortable to wear but reduce swelling.
- You feel worried about having another leg clot periodically on most days for about 3 months and very seldom thereafter.

Rating approach

Instructions

- To help people understand how good or bad a health state is (utility of a health state), we have drawn a scale
- On this scale, "Full health" is marked 100 and "dead" is marked 0
- We would like you to indicate on this scale how good or bad the marker health state is in your opinion
- Please consider the symptoms, time horizon, testing and treatment, and consequences described above and mark an arrow to indicate how good or bad the marker health state is
- Please write down below the number you marked on this scale



Using the VAS scale on the right, please rate the utility of the health state '**Severe DVT in the leg**', with a score of 100 representing a patient being in the "Full Health" and a score of 0 a patient being in "dead".

The number you marked is _____

Chapter 2: Appendix B - Health outcome descriptors developed for ASH VTE guidelines

Table B1: HODs developed for ASH VTE guidelines

Health Outcome Descriptor	Health Outcome Descriptor (continued)
1. Acne - Mild	66. Hospitalization Adults/Adolescents
2. Acne - Severe	67. Hospitalization Children
3. Acute Coronary Syndrome – Non-ST Elevation Myocardial Infarction (NSTEMI)	68. Inadequate Patient Medication Adherence
4. Acute Coronary Syndrome – ST Elevation Myocardial Infarction (STEMI)	69. Increased Duration of Hospitalization
5. Acute Kidney Injury	70. Infant Bleeding - Mild
6. Acute Liver Failure	71. Infant Bleeding - Severe
7. Acute Pain - Mild	72. Interacting Medications
8. Acute Pain - Moderate	73. Intracardiac Shunt (Glenn or Fontan) Thrombosis in a Child
9. Acute Pain - Severe	74. Irregular Menses
10. Adrenal Insufficiency	75. Ischemic Stroke - Mild
11. Allergic Reaction to Contrast Dye - Mild	76. Ischemic Stroke - Severe
12. Allergic Reaction to Contrast Dye - Moderate	77. IVC Filter Failure
13. Allergic Reaction to Contrast Dye - Severe	78. Limb Amputation
14. Any Diagnostic Test - False Negative	79. Low Time in Therapeutic INR Range; High INR Variability
15. Any Diagnostic Test - False Positive	80. Major Bleeding
16. Any Diagnostic Test - Inconclusive	81. Maternal Breast Irradiation
17. Any Diagnostic Test - True Negative	82. Maternal Plasma Drug Anticoagulation Level
18. Any Diagnostic Test - True Positive	83. Maternal Skin Reactions
19. Aortic Aneurysm	84. Menorrhagia
20. Bleeding during Oocyte Retrieval	85. Mesenteric Vein Thrombosis – Acute
21. Burden of Therapy - Injections in Pregnancy	86. Mesenteric Vein Thrombosis – Sub-Acute
22. Burden of Therapy – Diagnostic Procedure	87. Mild Anxiety
23. Central Venous Line Dysfunction in an Infant	88. Minor Bleeding
24. Cerebral Venous Thrombosis – Mild	89. Moderate to Severe Anxiety
25. Cerebral Venous Thrombosis – Severe	90. Multiple Organ Failure

26. Cesarean Wound Complication	91. Neonatal Bleeding - Mild
27. Chronic Kidney Disease	92. Neonatal Bleeding - Severe
28. Chronic Liver Failure	93. Osteoporosis
29. Chronic Pain - Mild	94. Peripheral Arterial Disease
30. Chronic Pain - Moderate	95. Permanent Disability or Dependency
31. Chronic Thrombotic Pulmonary Hypertension	96. Placental Abruption – Non-Severe
32. Cognitive Impairment - Mild	97. Placental Abruption – Severe
33. Cognitive Impairment - Severe	98. Portal Vein Thrombosis – Acute
34. Congenital Malformation	99. Portal Vein Thrombosis – Chronic
35. Contrast Induced Nephropathy	100. Portal Vein Thrombosis in a Child
36. Critical INR	101. Post-Thrombotic Syndrome
37. CVC-Related Thrombosis in Adolescents - Severe	102. Preeclampsia
38. CVC-Related Thrombosis in Infants	103. Pregnancy Loss
39. Deep Venous Thrombosis (DVT) in the Forearm – Mild	104. Preterm Delivery
40. Deep Venous Thrombosis (DVT) in the Forearm – Moderate	105. Preterm Labor
41. Deep Venous Thrombosis (DVT) in the Forearm – Severe	106. Prolonged Immobilization
42. Deep Venous Thrombosis (DVT) in the Lower Leg – Mild	107. Psychological Burden of Diagnostic Labels
43. Deep Venous Thrombosis (DVT) in the Lower Leg – Moderate	108. Pulmonary Embolism - Mild
44. Deep Venous Thrombosis (DVT) in the Lower Leg – Severe	109. Pulmonary Embolism - Moderate
45. Deep Venous Thrombosis (DVT) in the Upper Arm – Mild	110. Pulmonary Embolism - Severe
46. Deep Venous Thrombosis (DVT) in the Upper Arm – Moderate	111. Pulmonary Function
47. Deep Venous Thrombosis (DVT) in the Upper Arm – Severe	112. Quality of Life Impairment
48. Deep Venous Thrombosis (DVT) in the Upper Leg – Mild	113. Radiation Exposure
49. Deep Venous Thrombosis (DVT) in the Upper Leg – Moderate	114. Renal Vein Thrombosis in a Child - Bilateral
50. Deep Venous Thrombosis (DVT) in the Upper Leg – Severe	115. Renal Vein Thrombosis in a Child - Unilateral
51. Delay of Intervention	116. Reoperation
52. Dysmenorrhea	117. Retinal Vein Occlusion
53. Elevated Liver Values	118. Skin Complications from Compression Stockings
54. Emergency Room Visit	119. Small for Gestational Age
55. Fetal Radiation Exposure	120. Spinal Epidural Hematoma
56. Gastrointestinal Bleeding - Major	121. Splenomegaly
57. Gastrointestinal Bleeding - Minor	122. Systemic-to-Pulmonary Shunt Thrombosis in an Infant
58. Heart or Lung Transplantation	123. Unintended Pregnancy
59. Hemiplegia	124. Venous Ulcer

60. Hemorrhagic Stroke	125. Vulvar Hematoma
61. Heparin Skin Necrosis	126. Wound Hematoma
62. Heparin-Associated Bone Fractures due to Osteoporosis	127. Wound Infection
63. Heparin-Induced Thrombocytopenia (HIT)	
64. HIT Test – False Negative	
65. HIT Test – False Positive	

Chapter 2: Appendix C - Health outcome descriptor implementation in the guideline process

Figure C1: Outcome importance rating

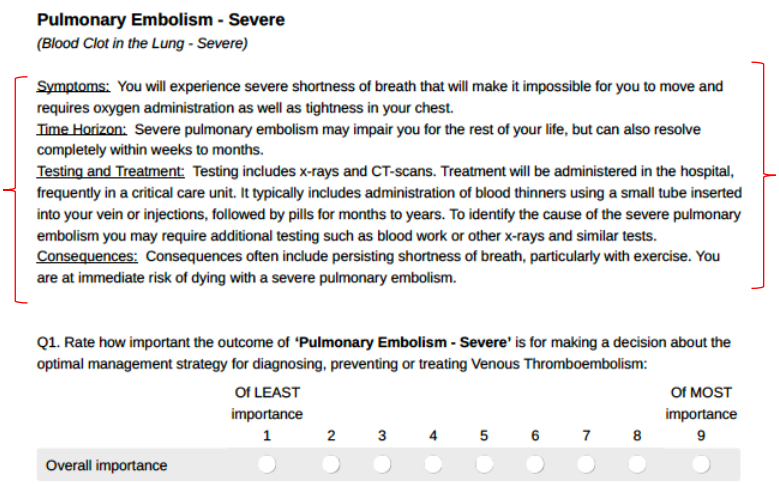


Figure C1 Caption: HODs were incorporated into outcome importance rating online surveys completed by panelists.

Figure C2: Outcome utility rating

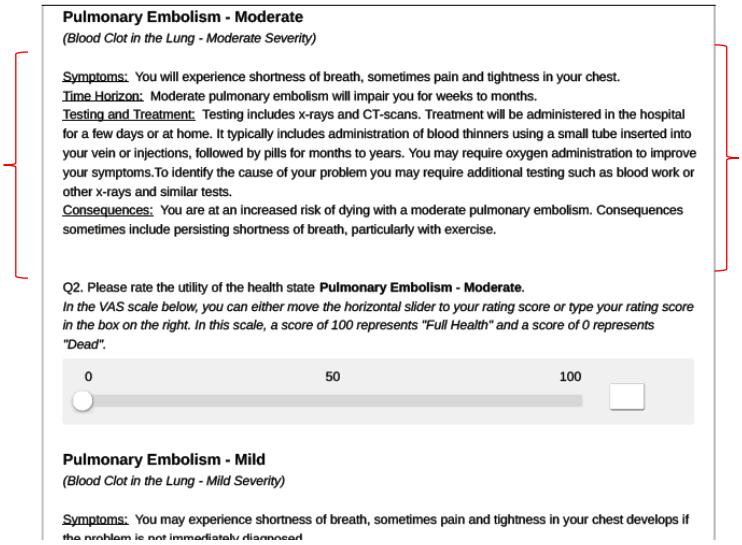


Figure C2 Caption: HODs were incorporated into outcome utility rating surveys completed by panelists.

Figure C3: Summary-of-findings tables

Summary of findings:

Any parenteral anticoagulation (UFH, LMWH or fondaparinux) compared to no parenteral anticoagulant in acutely ill medical patients for VTE prophylaxis

Patient or population: acutely ill medical patients for VTE prophylaxis
Setting: Inpatient
Intervention: any parenteral anticoagulation (UFH, LMWH or fondaparinux)
Comparison: no parenteral anticoagulant

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Com
	Risk with no parenteral anticoagulant	Risk with any parenteral anticoagulation (UFH, LMWH or fondaparinux)				
Mortality assessed with: all-cause mortality	69 per 1,000	**67 per 1,000** (63 to 72)	**RR 0.97** (0.91 to 1.04)	49002 (21 RCTs) 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20	⨁⨁◯◯ LOW [a,b]	
Symptomatic Pulmonary Embolism - representing the moderate marker state assessed with: symptomatic PE	**Study population**		**RR 0.59** (0.45 to 0.78)	25687 (13 RCTs) [2,5,8,11,12,13,14,15,17,19,20,21,22]	⨁⨁⨁◯ MODERATE [a,c]	
	10 per 1,000	**6 per 1,000** (5 to 8)				
	Low					
	0 per 1,000 [d]	**0 per 1,000** (0 to 0)				
Proximal Deep Vein Thrombosis - representing the moderate marker state assessed with: symptomatic proximal DVT	**Study population**		**RR 0.28** (0.06 to 1.37) [e]	3706 (1 RCT) [2]	⨁⨁⨁◯ MODERATE [b]	
	4 per 1,000	**1 per 1,000** (0 to 5)				
	Low					
	1 per 1,000 [d]	**0 per 1,000** (0 to 1)				
Distal Deep Vein Thrombosis - representing the moderate marker state assessed with: symptomatic distal DVT	**Study population**		**RR 0.75** (0.17 to 3.34) [f]	3706 (1 RCT) [2]	⨁⨁⨁◯ MODERATE [b]	
	2 per 1,000	**2 per 1,000** (0 to 7)				
	Moderate					
	14 per 1,000	**11 per 1,000** (2 to 48)				
Major Bleeding	7 per 1,000	**10 per 1,000** (6 to 19)	**RR 1.48** (0.81 to 2.71)	30761 (16 RCTs) 3,5,6,7,8,9,10,11,12,13,14,15,17,19,20,22	⨁⨁◯◯ LOW [a,b]	
Gastrointestinal Bleeding	31 per 1,000	**82 per 1,000** (11 to 589)	**RR 2.61** (0.36 to 18.86)	185 (2 RCTs) [23,24]	⨁⨁◯◯ LOW [a,b]	
Heparin-Induced Thrombocytopenia	2 per 1,000	**2 per 1,000** (1 to 4)	**RR 0.95** (0.47 to 1.92)	12577 (3 RCTs) [1,2,3]	⨁⨁⨁◯ MODERATE [b]	

Figure C3 Caption: Outcomes defined by HODs that were selected as critical for decision-making based on the panel's prioritization exercise were included for evidence syntheses and presented in summary-of-findings tables. When reviewing the estimates of effects from systematic reviews, panelists would see the specific HOD considered by the panel for decision-making according to the label (e.g. "pulmonary embolism – representing the moderate marker state"). For example, in this guideline, the panel considered that the outcome of pulmonary embolism occurring in this population would typically be that of moderate severity, as opposed to mild or severe severity, and prioritized this HOD.

CHAPTER 3. NEW METHODS FACILITATED THE PROCESS OF PRIORITIZING QUESTIONS AND HEALTH OUTCOMES IN GUIDELINE DEVELOPMENT

New methods facilitated the process of prioritizing questions and health outcomes in guideline development

Authors
Wojtek Wiercioch[1,2]
Yuan Zhang[1,2]
Robby Nieuwlaat[1,2]
Pablo Alonso-Coello[3]
Philipp Dahm[4]
Alfonso Iorio[2,5]
Veena Manja[6,7]
Reem A. Mustafa[2,8]
Ignacio Neumann[9]
Thomas L. Ortel[10]
Bram Rochwerg[2,5]
Nancy Santesso[1,2]
Sara K.Vesely[11]
Elie A. Akl[2,12]
Holger J. Schünemann[1,2,5]

Author Affiliations:

1. Michael G. DeGroote Cochrane Canada & McMaster GRADE Centres; Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
2. Department of Health Research Methods, Evidence, and Impact, McMaster University, Canada
3. Iberoamerican Cochrane Center, Biomedical Research Institute Sant Pau (IIB Sant Pau-CIBERESP), Barcelona, Spain
4. Minneapolis VAMC, Urology Section and University of Minnesota Department of Urology, Minneapolis, MN
5. Department of Medicine, McMaster University, Hamilton, Canada
6. Department of Surgery, University of California Davis, CA, USA
7. Department of Medicine, Veterans Affairs Northern California Health Care System, Mather, CA
8. Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, Kansas City, KS, USA
9. Department of Internal Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile
10. Departments of Medicine and Pathology, Duke University Medical Center, Durham, NC
11. Department of Biostatistics and Epidemiology, Hudson College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA
12. Department of Internal Medicine, American University of Beirut, Lebanon

Corresponding author:
Holger J. Schünemann
Department of Health Research Methods, Evidence and Impact
McMaster University, HSC-2C
1280 Main St West; Hamilton, ON, L8N 3Z5, Canada
e-mail: schuneh@mcmaster.ca

Abstract

Background: Health guideline development requires sequential prioritization of the guideline topic, questions, and health outcomes. In this paper we report on new approaches for prioritizing questions and health outcomes in guideline development.

Methods: Ten guideline panels on venous thromboembolism (VTE) rated potential guideline questions on a 9-point scale according to their overall importance and 6 criteria: common in practice, uncertainty in practice, variation in practice, new evidence available, cost consequences, not previously addressed. We randomized panelists to rate one potential question with and without the 6 criteria. Panelists rated importance of outcomes, defined with health outcome descriptors (HODs), using a 9-point scale, and health utility of outcomes on a visual analogue scale.

Results: Of 469 potential questions identified, 72.5% were rated as important but not of high priority and 25.4% as high priority. Each criterion was significantly associated with the overall importance rating. The means for the overall importance ratings were 5.96 (SD 2.38) and 6.53 (SD 2.45) (p = 0.25) for those randomized to rate questions with and without the criteria, respectively. The mean importance rating for 121 outcomes was 6.01 (SD 1.25), with 35.5% rated as critical for decision-making. Panelists provided health utility ratings for 127 outcomes, with a mean utility rating of 0.56 (SD 0.19) and minimum mean and maximum mean utility ratings of 0.12 and 0.91, respectively.

Conclusions: Our structured guideline development steps provided information to help explain question importance ratings, facilitate panels' outcome prioritization, and information for decision-making in guideline development.

Keywords: clinical practice guidelines, expert panels, healthcare question prioritization, health outcome importance, health outcome utility

What is new:

- We present a survey approach that informed question prioritization by guideline panels with rating of overall importance of questions and for six additional criteria that were found to be associated with the overall importance.
- Panels' question importance ratings informed discussions and consensus about which questions to address in a guideline. An overall importance rating may be sufficient to inform these deliberations in most scenarios. We found that the approach reduced and classified an initial list of potential questions in ten guidelines by 75% to those deemed of high priority.

- Additional explicit rating criteria may provide insight as to why panel members rate specific questions as important to inform panel discussions, but did not impact overall importance ratings.
- Using a survey, which incorporated health outcome descriptors to define health outcomes, provided an approach for panels to rate the importance of health outcomes to inform decisions about which outcomes to consider in formulating recommendations.
- Using a survey to rate the health utility of outcomes with guideline panels provided an approach to collect information that may inform panels' decisions about the relative value placed on health outcomes, but it will need to be validated in future work.

Abbreviations

ASH – American Society of Hematology
DVT – Deep vein thrombosis
GIN – Guidelines International Network
GRADE – Grading of Recommendations Assessment, Development and Evaluation
HOD – Health outcome descriptor
PE – Pulmonary embolism
VAS – Visual analogue scale
VTE – Venous Thromboembolism

1. Introduction

Health care guideline development requires staged and sequential prioritization exercises. This includes the prioritization of the guideline topic, the prioritization of the questions that will be answered by the recommendations, and finally the health outcomes that describe the benefits and harms affected by the alternative interventions being considered.[1-7] These critical steps ultimately determine how relevant and useful the product will be for end users including clinicians, patients, and policymakers. While the guideline topic is often decided by the organization that produces a guideline, the formulation and prioritization of questions and selection and prioritization of outcomes typically involves the guideline development group, also called the guideline panel ("panels").

Guideline developers and expert panels are tasked with prioritizing relevant, timely, and important questions for which they will formulate recommendations. This process involves the selection, for each potential question, of the population and interventions of interest. Panels often face the challenge of narrowing down a long list of possible questions, to a limited list that can be feasibly addressed within the time and resource constraints of the specific guideline development project.

Once the key questions are selected, guideline panels face another challenge. They have to identify the critical (or at least important) health outcomes for which to synthesize the research evidence and upon which to formulate their recommendations.[1] Panels must weigh the impact of interventions on patient-important health outcomes to

determine the balance of health benefits and harms. Information on what are the important health outcomes, and how important they are to patients, may in some cases, be available from the published literature.[2,6-9]

Systematic reviews can provide the necessary information on how patients value specific outcomes (e.g. through reporting of health values or utilities).[10-13] However, these evidence syntheses may not exist for the topic of interest, and when they do, they may not identify information relevant to the specific outcomes of interests. Alternatively, panel members, including particular patient representatives and clinical experts ,may act as a proxy for obtaining information on outcome importance or help with determining which outcomes to focus on in new evidence syntheses.[6,14-18]

For both the prioritization of questions and the outcomes related to those questions, guideline developers typically use survey methods,[19,20] informal discussion and consensus, or formal consensus methods (e.g. Delphi approach), each with specific advantages and limitations.[1,4,16-18,21] Indeed, optimal methods for carrying out prioritization in guidelines remains an underdeveloped area of guideline development and one in need or further research.[22] We addressed this question in a recent project which involved ten guideline panels and that developed over 250 recommendations for the American Society of Hematology (ASH).[23-29] This article describes the methodology and findings of the approach to prioritize questions, rate the importance of health outcomes and derive health utility values with guideline panels. It follows our

companion publication in which we described the development of health outcome descriptors (HODs) implemented in these approaches.[30]

2. Methods

For the development of venous thromboembolism (VTE) guidelines, ASH formed ten multidisciplinary panels, each with memberships consisting of 11-17 panelists(see Appendix A for the guideline topics).[31,32] Clinical and methodology co-chairs lead the panels which included one to two patient representatives in addition to other clinical and methodology experts. The overall guideline process was based on the Guideline International Network (G-I-N)-McMaster Guideline Development Checklist and the GRADE approach (Grading of Recommendations Assessment, Development and Evaluation) and was guided by the McMaster University GRADE Centre and ASH.[3,33,34]

2.1 Guideline question prioritization

After an initial step of brainstorming by the panels to generate an exhaustive list of potential questions for their individual guidelines, we developed and administered online question prioritization surveys using SurveyMonkey.[35] We formatted potential guideline questions using the PICO (population, intervention, comparison, outcome) framework and standardized structure for questions about management or diagnosis (e.g. Should intervention/strategy A versus intervention/strategy B be used in patients with X condition/characteristic for the prevention of venous thromboembolism).[6]

The prioritization surveys asked panel members to rate each question according to six criteria and its overall importance, using a 9-point scale (1: least important; 9:

most important). A sample question prioritization survey is provided in Appendix B. Panel members rated each question as to whether it was one: (i.) that commonly arises in practice; (ii.) for which there is uncertainty in practice regarding how to manage patients; (iii.) for which there is new research evidence to consider; (iv.) that is associated with variation in practice; (v.) that has important consequences for, or is associated with, high resource use or costs; and (vi.) that has not been previously or sufficiently addressed (e.g. in previous guidelines). We based these six criteria on prior literature and input from guideline methodologists.[1,4,21,36-38] We hypothesized that these criteria would be predictors of the overall importance of a potential guideline question and would assist guideline panelists in judging the importance of questions. To explore if explicit rating with the 6 criteria leads to different results because panelists are reminded of what to consider, we also randomized panel members to rate the first potential guideline question in the prioritization survey either with or without rating the additional 6 criteria. Regardless of which arm panel members were randomized to, all rated the overall importance of the question. Panelists randomized to rate the first question with the overall importance only, were then shown the question again and asked to rate the overall importance as well as the 6 criteria. We used the built-in randomization sequence within SurveyMonkey.

We summarized the panel members' mean and median ratings and presented them to the panels to guide discussion and reach consensus on a final list of approximately 20 questions per guideline panel. When presenting the findings, we used

the mean rating to colour-code and categorize the questions as being of high priority to address in the guideline (rating of 7 to 9), an important question but not of high priority (rating of 4 to 6), or of low priority (rating of 1 to 3) (see Appendix C showing an example of a summary with the colour coding approach). We decided *a priori* that the panels should first base their decisions on the overall importance ratings and use the additional criteria ratings to choose between multiple questions receiving the same or similar overall importance ratings. In addition, we used the additional criteria to interpret why a question may have received higher or lower overall importance rating.

2.2 Outcome importance rating

We used online surveys for the rating of importance of health outcomes. Rating of outcome importance followed an initial step where panels brainstormed a list of health outcomes considered as relevant for the questions. To create common definitions for the outcomes in the rating exercise, we developed a health outcome descriptor (HOD) for each potential outcome that included a description of the symptoms, time horizon, testing and treatment, and consequences (see examples in Figure 1). We describe the concepts and development of the HODs in detail in the companion publication.[30]

All brainstormed outcomes were compiled into one list and the importance rating surveys used between the ten guideline panels differed only in the order of outcomes presented according to the topic. The surveys were structured into 3 sections: 1) outcomes brainstormed by the specific panel, 2) outcomes brainstormed by other

panels, but of possible relevance to the panel's topic, 3) outcomes brainstormed by other panels, and unlikely to be of relevance to the panel's topic (see an example of the survey in Appendix D). Panelists rated the importance of each outcome on a 9-point scale, with a rating of 7 to 9 indicating the outcome as being critical for decision-making, 4 to 6 as important but not critical, and 1 to 3 of limited or no importance.[6] Individual panelists serving on multiple panels were asked to complete the survey only once.

2.3 Outcome utility rating

In a final step, we administered an online survey for panel members to rate the utility of health outcomes; an approach derived from health economics to value outcomes.[39] Panel members rated the utility of outcomes on a 0 to 100 visual analogue scale (VAS), with a rating of 0 indicating the state of being dead and 100 indicating the state of full health (see Figure 2 and Appendix E for an example utility rating survey). As with the outcome importance rating survey, HODs were provided for each outcome to aid in this exercise. Panelists from all ten guidelines rated the utility of outcomes to obtain representative ratings and allow for comparisons between panels' ratings. The panels' outcome utility ratings were used in the evidence-to-decision step to inform panels' decisions about the relative value of outcomes when weighing health benefits and harms during the formulation of recommendations. The health utilities derived from the exercise were used in conjunction with information identified in a systematic review of patients' values and preferences.[13]

2.4 Analysis

For descriptive analysis of question prioritization, outcome importance and utility ratings we calculated means and standard deviations. We used an independent samples *t*-test to compare the mean overall question importance ratings of panel members randomized to rate the first question in their surveys with and without the additional 6 criteria. We used a mixed effects linear regression model to determine whether each of the 6 criteria for question prioritization were significant predictors of overall question importance. The mixed effects model with panel member treated as a random effect and the 6 predictor variables treated as fixed effects was selected to account for variance within-subjects and between-subjects as well as the unbalanced design, with panel members rating different numbers of potential questions per guideline. To classify the outcome importance ratings into the 3 categories of the 9-point scale we rounded the mean ratings to the nearest full digit. We converted the utility ratings from the VAS to a 0 to 1 scale. We considered surveys with more than 80% missing data incomplete and removed them from the data sets. We completed analyses using IBM SPSS Statistics 19 and Microsoft Excel.

3. Results

3.1 Guideline question prioritization

One-hundred-and-fourteen of 131 (87%) panel members across the ten guidelines completed the question prioritization surveys, of which 97 (85%) provided complete data. Across the ten guidelines, the number of questions rated in the surveys

ranged from 19 to 112 (median 38), with a mean completion time ranging from 28 to 62 minutes. Overall, this gives a mean time of 65 seconds (standard deviation (SD) 31 seconds) required to rate all 6 criteria and the overall importance (total of 7 ratings) for each question (see Appendix A)

Across all guidelines and questions, the mean overall importance rating was 5.75 (SD 2.1) on the 9-point scale (Table 1). Of note, across the 469 questions rated in the ten guidelines, 119 (25.4%) of the questions received a mean overall importance rating between 7 to 9 (high priority to address in the guideline), 340 (72.5%) received a rating between 4 to 6 (important question but not of high priority) and ten (2.1%) received a rating of 1 to 3 (low priority) (Figure 3).

The mean ratings for each of the 6 criteria ranged from 4.48 to 6.11 and each of the criteria had a positively linear relationship with the overall importance rating (Table 1). Applying the linear mixed effects model, each of the individual criteria ratings were shown to be significant predictors of overall importance in a univariable analysis (β ranging from 0.64 to 0.77, $p<0.01$ for all criteria) (see Appendix E for the model). When included in an adjusted analysis, each criterion remained a significant predictor of overall importance (β 0.12 to 0.29, $p<0.01$ for all criteria) (Table 1). The two criteria with the strongest relationship with overall importance were: question being a common one in practice and not previously addressed.

In the randomized trial, 43 panelists were randomized to rate the first question in the survey only with the overall importance rating, and 50 were assigned to rate the

question with the 6 criteria and the overall importance. The respective means for the overall importance ratings were 6.53 (SD 2.45) and 5.96 (SD 2.38) (p = 0.25). When rating the question again with the six criteria, 13 of the 43 panelists assigned a lower overall importance rating, 10 assigned a higher rating (mean change in rating of -0.41, SD 1.69), and 20 had no change.

3.2 Outcome importance rating

The outcome importance rating survey included 121 HODs for 99 health outcomes (more than one HOD was described for outcomes with different levels of severity) and was completed by 111 of 118 (94%) panel members. The mean importance rating based on rating by all panels was 6.01 (SD 1.25), with a minimum mean rating of 3.19 (for mild allergic reaction to contrast dye) and a maximum mean rating of 8.76 (for severe pulmonary embolism) on the 9-point scale (Table 2). Of the 121 HODs, 1 (0.8%) received a mean rating of ≤3 (limited or no importance), 77 (63.6%) received a mean rating of 4 to 6 (important but not critical), and 43 (35.5%) were rated ≥7 (critical for decision making). There was good agreement in the importance rating between the panels and an aggregate measure of importance. Out of 1210 panel mean ratings for the outcomes, 990 (82%) were classified into the same importance category as determined by the grand mean across the 10 panels (Table 2).

3.3 Outcome utility rating

The outcome utility rating survey included 127 HODs for 104 health outcomes, as 5 additional outcomes were added by one guideline panel based on feedback received

the importance rating survey. The survey was completed by 79 of 118 (67%) panel members, however, 9 panelists used the VAS scale in reverse, consistently assigning higher utility rating to more severe outcomes, and were not included in the pooled estimates for a total of 70 valid ratings. The mean utility rating was 0.56 (SD 0.19), with a minimum mean rating of 0.12 (for hemorrhagic stroke) and a maximum mean rating of 0.91 (for true negative diagnostic test result) on the VAS (Table 2). There were 6 outcomes rated by the panels for which health utility values were also reported in studies eliciting health utilities from patients or the general population, which were identified in the systematic review of patients' values and preferences (see Table 3). There was only some overlap in the utility ratings of the panels and ranges of the utility ratings reported in the literature. This is likely due to variability in methods used for eliciting utilities (e.g. time trade-off, standard gamble, VAS), as was also noted across the individual studies included in the systematic review.

4. Discussion

As part of a guideline development project involving ten guideline panels, we addressed two prominent challenges of guideline development for which there is little empirical evidence. We evaluated the impact of providing a structured approach for panels: 1) to guide their prioritization of potential guideline questions and, thus, the recommendations they would develop, and 2) to serve as a proxy for information about the importance of health outcomes on which to base decision-making during the formulation of recommendations. Two recent systematic reviews by El-Harakeh and

colleagues,[22,40] identified, respectively, approaches and exercises that focused on prioritizing guideline topics but none on prioritizing recommendation questions or outcomes. To our knowledge, this is the first study reporting detailed methods and empirical evidence for this critical step in guideline development.

Providing detailed criteria for rating the importance of potential guideline questions can serve to inform panels' deliberations about why specific questions are deemed important, and facilitate decisions about choosing questions for the guideline topic. Requesting ratings for additional criteria was a feasible approach, even with rating a large number of questions, and we achieved a survey completion rate of 94% with our panelists. While our findings generally suggest that an overall question importance rating may be sufficient for most scenarios, obtaining additional ratings could also serve to provide information that may help resolve 'ties' that often occur with simple overall importance ratings; an aspect that could be evaluated in future research. In the second step of prioritizing health outcomes with the use of HODs, panel members were able to rate the importance of health outcomes across a range of topics to select outcomes considered critical for judging health benefits and harms. We also provide utility estimates from the panels for many HODs that have not been previously evaluated. From a practical standpoint, we demonstrate that structured prioritization of guideline questions and health outcomes is a feasible approach to inform guideline panels' decision-making.

4.1 Strengths

Our study has several strengths. We planned this research prior to developing the guidelines and embedded it in a major guideline development project with ten panels. Inclusion of the research in a real guideline development exercise allows inferences about its feasibility and increases the validity and applicability of our approaches. We enrolled multiple panels and used randomization in our study design to explore whether simple methods for question prioritization are exchangeable with a more detailed approach. We also conducted the outcome importance and utility rating exercises with a large sample of panel members including patient representatives, and applied HODs to facilitate a common understanding of health outcomes during the rating. As we conducted the utility rating exercise in parallel with a systematic review on utilities for VTE-related outcomes as part of the guideline development project, we were able to compare the panels' utility ratings with those reported in the literature.

4.2 Limitations

For the question prioritization rating exercise, we did not explore whether the overall importance rating or the criterion ratings predicted the final inclusion of a question in a guideline, which is a topic for immediate future research. For the outcome brainstorming and rating exercises, panels rated the importance of outcomes across guideline questions. However, using HODs which focus on consequences for patients, the importance and interpretation of outcomes is not expected to vary by guideline question. Additionally, the questions across the VTE guidelines were similar in context and the large number of panelists conducting the brainstorming and rating exercises

should have ensured that no important outcomes were missed. For the utility rating exercise, despite detailed instructions in the online survey and providing the diagram of the VAS (Figure 2), 9 of 88 respondents used the VAS in reverse and were excluded from the analysis. Therefore, additional guidance and instruction is needed for panelists to apply the VAS correctly to rate outcome utilities. Furthermore, this can be addressed by automated internal consistency checks using online ratings, e.g. by warnings to respondents when they rate a mild HOD as more severe than a severe HOD. A limitation of the utility rating approach that requires further evaluation is whether ratings by guideline panels are a suitable proxy for utility ratings by patients and the general population that would experience the outcomes.

4.3 Implications for Guideline Developers

The approaches for question prioritization and health outcome prioritization and utility rating described in this study are feasible approaches to engage panels in providing information that is could help inform the formulation of recommendations. If resources and time allow, this approach could potentially be used with external stakeholders (e.g. patient advocacy group, medical specialty association membership) for broader input to inform these steps of the guideline development process. Obtaining data about the importance and utility of health outcomes provides information that is necessary for decision-making regarding recommendations but may not be reported in the literature.

4.4 Next Steps and Future Research

Our regression analysis suggests that each criterion in our list of six is associated with prioritizing questions in VTE guidelines. Future research should examine the application of the six criteria for question prioritization to different guideline topics. While we evaluated 6 criteria that we were aware of, with the aim to identify the ones that are the most appropriate when prioritizing questions, additional work could explore if other factors influence prioritization.[22,40] The regression analysis also suggests that additional factors exist. Qualitative evaluation of how panels use the question rating information to make final decisions about question inclusion would provide further guidance on how panels can prioritize the key health questions for their guidelines. For prioritization and utility rating of health outcomes, comparisons between ratings by expert panels consisting of clinicians, patient representatives and other decision-makers to that of external stakeholder groups would be of interest to validate using expert panels as a proxy for this information.

4.5 Conclusions

Typically, the selection of questions for inclusion in guidelines has been unstructured, lacking transparency, and without much specific guidance on what factors to consider. The detailed approach we present for question prioritization provides information that can assist panels with selecting questions for which they will formulate recommendations. Health outcome prioritization using HODs facilitates panels' understanding of key information on the importance and value of health outcomes and informs their deliberations during the formulation of recommendations.

Funding/support: The development of the VTE guidelines was funded by the American Society of Hematology.

Role of the funding source: The funder supported the guideline development work, including conduct of the panel survey exercises that provided data for this research study, but did not have any role in determining methodology and reporting.

Financial Disclosures / Conflict of interest:

Elie Akl has conducted methodological research on the topic of prioritization of guidelines

T.L.O has served as a consultant for Instrumentation Laboratory, and his institution has received research support on his behalf from Instrumentation Laboratory, Siemens, Stago, NIH, CDC, and PCORI.

All other authors declare no competing interests.

Contributors: HJS was the principal investigator. HJS, RN, WW conceptualized the study and all co-authors provided feedback on the approaches the research study. WW was responsible for carrying out the data analysis. WW drafted the manuscript with critical revisions and writing contributions from HJS. All of the authors revised the manuscript critically for important intellectual content and approved the final version submitted for publication.

Acknowledgements: The authors thank the ASH VTE guidelines panel members for participating in the question and outcome rating exercises as part of the guideline development project.

References

1. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 2. Priority setting. *Health Res Policy Syst.* 2006;4:14.
2. Schunemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. *Health Res Policy Syst.* 2006;4:18.
3. Schunemann 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):E123-142.
4. Atkins D, Perez-Padilla R, Macnee W, et al. Priority setting in guideline development: article 2 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proceedings of the American Thoracic Society.* 2012;9(5):225-228.
5. Eccles MP, Grimshaw JM, Shekelle P, Schunemann HJ, Woolf S. Developing clinical practice guidelines: target audiences, identifying topics for guidelines, guideline group composition and functioning and conflicts of interest. *Implement Sci.* 2012;7:60.
6. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011;64(4):395-400.
7. Kelson M, Akl EA, Bastian H, et al. Integrating values and consumer involvement in guidelines with the patient at the center: article 8 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. *Proceedings of the American Thoracic Society.* 2012;9(5):262-268.
8. Zhang Y, Coello PA, Guyatt GH, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences-inconsistency, imprecision, and other domains. *J Clin Epidemiol.* 2019;111:83-93.
9. Zhang Y, Alonso-Coello P, Guyatt GH, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-Risk of bias and indirectness. *J Clin Epidemiol.* 2019;111:94-104.
10. Tarabay R, El Rassi R, Dakik A, et al. Knowledge, attitudes, beliefs, values, preferences, and feasibility in relation to the use of injection safety devices in healthcare settings: a systematic review. *Health Qual Life Outcomes.* 2016;14(1):102.
11. 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.
12. Zhang Y, Morgan RL, Alonso-Coello P, et al. A systematic review of how patients value COPD outcomes. *Eur Respir J.* 2018;52(1).

13. Etxeandia-Ikobaltzeta I, Zhang Y, Brundisini F, et al. Patient values and preferences regarding VTE disease: a systematic review to inform American Society of Hematology guidelines. *Blood Advances.* 2020;4(5):953-968.
14. Falzon D, Jaramillo E, Schunemann HJ, et al. WHO guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update. *Eur Respir J.* 2011;38(3):516-528.
15. Schünemann HJ BJ, Oxman AD, Guyatt GH. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. . In: Schünemann HJ BJ, Oxman AD, Guyatt GH, ed. 2015 ed.: GRADEpro; 2015: http://gdt.guidelinedevelopment.org/central_prod/_design/client/handbook/handbook.html. Accessed March 20 2016.
16. Yepes-Nunez JJ, Fiocchi A, Pawankar R, et al. World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Vitamin D. *The World Allergy Organization journal.* 2016;9:17.
17. Cuello-Garcia CA, Fiocchi A, Pawankar R, et al. World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Prebiotics. *The World Allergy Organization journal.* 2016;9:10.
18. Fiocchi A, Pawankar R, Cuello-Garcia C, et al. World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Probiotics. *The World Allergy Organization journal.* 2015;8(1):4.
19. Ajuebor O, Cometto G, Boniol M, Akl EA. Stakeholders' perceptions of policy options to support the integration of community health workers in health systems. *Human Resources for Health.* 2019;17(1):13.
20. Darzi AJ, Officer A, Abualghaib O, Akl EA. Stakeholders' perceptions of rehabilitation services for individuals living with disability: a survey study. *Health Qual Life Outcomes.* 2016;14(1):2.
21. Reveiz L, Tellez DR, Castillo JS, et al. Prioritization strategies in clinical practice guidelines development: a pilot study. *Health Research Policy and Systems.* 2010;8:7-7.
22. El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. *BMC Health Services Research.* 2019;19(1):692.
23. Lim W, Le Gal G, Bates SM, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. *Blood Advances.* 2018;2(22):3226-3256.
24. Cuker A, Arepally GM, Chong BH, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. *Blood Advances.* 2018;2(22):3360-3392.
25. Witt DM, Nieuwlaat R, Clark NP, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: optimal management of anticoagulation therapy. *Blood Advances.* 2018;2(22):3257-3291.

26. Schünemann HJ, Cushman M, Burnett AE, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. *Blood Advances.* 2018;2(22):3198-3225.
27. Monagle P, Cuello CA, Augustine C, et al. American Society of Hematology 2018 Guidelines for management of venous thromboembolism: treatment of pediatric venous thromboembolism. *Blood Advances.* 2018;2(22):3292-3316.
28. Bates SM, Rajasekhar A, Middeldorp S, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: venous thromboembolism in the context of pregnancy. *Blood Advances.* 2018;2(22):3317-3359.
29. Anderson DR, Morgano GP, Bennett C, et al. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. *Blood Advances.* 2019;3(23):3898-3944.
30. Wiercioch W, Nieuwlaat R, Dahm P, et al. Development and Application of Health Outcome Descriptors in the Production of Practice Guidelines. *Journal of clinical epidemiology.* 2020;Submitted.
31. Wiercioch W, Nieuwlaat R, Akl EA, et al. Methodology for the American Society of Hematology VTE guidelines: current best practice, innovations, and experiences. *Blood advances.* 2020;4(10):2351-2365.
32. Izcovich A, Cuker A, Kunkle R, et al. A user guide to the American Society of Hematology clinical practice guidelines. *Blood Advances.* 2020;4(9):2095-2110.
33. 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.
34. 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.
35. SurveyMonkey Inc. www.surveymonkey.com. Accessed March 04, 2020.
36. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Clinical guidelines: developing guidelines. *BMJ (Clinical research ed).* 1999;318(7183):593-596.
37. Institute of Medicine. *Setting Priorities for Clinical Practice Guidelines.* Washington, DC: The National Academies Press; 1995.
38. Lavis JN, Oxman AD, Lewin S, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 3: Setting priorities for supporting evidence-informed policymaking. *Health Research Policy and Systems.* 2009;7 Suppl 1(Suppl 1):S3.
39. Schunemann HJ, Armstrong D, Degl'innocenti A, et al. A randomized multicenter trial to evaluate simple utility elicitation techniques in patients with gastroesophageal reflux disease. *Med Care.* 2004;42(11):1132-1142.

40. El-Harakeh A, Lotfi T, Ahmad A, et al. The implementation of prioritization exercises in the development and update of health practice guidelines: A scoping review. *PLoS One.* 2020;15(3):e0229249.

Chapter 3: Tables

Table 1: Panelists' question ratings on the 9-point scale and estimates of effects from regression model

Question Rating Criterion	Panelists' Rating	Estimate of Effects
	Mean (SD)	β (95% CI)
Overall Importance	5.75 (2.1)	
Common Question in Practice	6.11 (2.3)	0.29 (0.24 to 0.34)
Uncertainty in Practice	5.97 (2.3)	0.12 (0.09 to 0.15)
New Evidence Available	4.48 (2.4)	0.14 (0.11 to 0.17)
Variation in Practice	5.96 (2.2)	0.15 (0.11 to 0.19)
Consequences for Cost	5.97 (2.3)	0.12 (0.10 to 0.15)
Not Previously Addressed	5.81 (2.2)	0.20 (0.16 to 0.24)

Table 1 Caption: Panelists' ratings on the 9-point scale (1: least important; 9: most important) across guideline questions and estimates of effects from regression analysis. The β value gives the effect of a one-unit increase in rating of a criterion on the change in overall importance on the 9-point scale, holding the remaining criteria constant.

Table 2: Outcome Importance and Utility Ratings

Outcome/Health Outcome Descriptor	Importance Rating – all panels (n=113)	Importance Rating – VTE in Medical Patients (n=12)	Importance Rating – VTE in Surgical Patients (n=13)	Importance Rating – VTE Treatment (n=12)	Importance Rating – Optimal Management (n=10)	Importance Rating – HIT (n=11)	Importance Rating – Thrombophilia (n=7)	Importance Rating – VTE in Cancer (n=15)	Importance Rating – Pediatric VTE (n=13)	Importance Rating – VTE in Pregnancy (n=10)	Importance Rating – Diagnosis of VTE (n=8)	Utility Rating – all panels (n=70)
	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD) Category	Mean (SD)
Acute Coronary Syndrome – Non-ST Elevation Myocardial Infarction (NSTEMI)	6.9 (0.52) Critical	6.6 (2.01) Critical	7 (1.28) Critical	6.2 (1.72) Important	7.4 (1.2) Critical	7.5 (0.89) Critical	7 (1.31) Critical	6.7 (1.74) Critical	6 (2.63) Important	7.7 (1.19) Critical	6.9 (0.93) Critical	0.38 (0.2)
Acute Coronary Syndrome – ST Elevation Myocardial Infarction (STEMI)	7.4 (0.54) Critical	7.2 (1.85) Critical	7.7 (1.42) Critical	6.9 (1.98) Critical	7.6 (1.02) Critical	8.1 (0.94) Critical	7.6 (0.49) Critical	7.1 (1.78) Critical	6.4 (2.79) Important	8.3 (0.9) Critical	7.1 (1.45) Critical	0.31 (0.19)
Acute Kidney Injury	6.5 (0.35) Important	6.8 (0.99) Critical	6.1 (2.02) Important	5.9 (1.66) Important	6.9 (0.83) Critical	6.5 (1.5) Important	6.4 (1.18) Important	6.9 (1.02) Critical	5.9 (1.73) Important	6.6 (2.31) Critical	6.6 (0.99) Critical	0.43 (0.19)
Acute Liver Failure	7.1 (0.46) Critical	7.5 (1.19) Critical	7.4 (1.87) Critical	6.5 (1.94) Critical	7.6 (1.11) Critical	7.3 (1.96) Critical	7 (1.6) Critical	7.5 (1.15) Critical	6.1 (2.46) Important	7.1 (2.51) Critical	7.3 (0.66) Critical	0.25 (0.16)
Acute Pain - Mild	3.7 (0.49) Important	3.3 (1.48) Not Important	3.4 (1.87) Not Important	3.4 (1.32) Not Important	4.5 (1.69) Important	4.2 (1.34) Important	3.3 (1.67) Not Important	4.6 (1.2) Important	3.6 (1.27) Important	3.4 (1.5) Not Important	3.9 (1.62) Important	0.78 (0.14)
Acute Pain - Moderate	5.5 (0.46) Important	5.1 (1.44) Important	5.4 (1.92) Important	5.5 (1.44) Important	6.4 (1.11) Important	6 (1.04) Important	5.1 (1.73) Important	6 (1.1) Important	4.8 (0.86) Important	5.3 (1.42) Important	5.4 (1.41) Important	0.55 (0.18)
Acute Pain - Severe	6.8 (0.63) Critical	6.7 (1.7) Critical	6.5 (2.23) Important	7 (1.41) Critical	7.7 (1.19) Critical	7.4 (1.07) Critical	6 (1.77) Important	7.5 (1.09) Critical	5.6 (1.33) Important	6.6 (1.85) Critical	7.1 (1.17) Critical	0.35 (0.19)
Adrenal Insufficiency	5.1 (0.56) Important	4.4 (1.98) Important	4.9 (2.11) Important	4.3 (1.7) Important	6.1 (1.64) Important	5.6 (1.82) Important	4.9 (1.73) Important	4.5 (2.06) Important	5.2 (1.96) Important	5 (3.13) Important	5.8 (1.3) Important	0.59 (0.2)
Allergic Reaction to Contrast Dye - Mild	3.2 (0.44) Not Important	3 (1.6) Not Important	2.8 (1.4) Not Important	3 (1.68) Not Important	4.4 (1.8) Important	3.3 (1.54) Not Important	2.9 (1.88) Not Important	3.5 (1.54) Not Important	3.1 (1.11) Not Important	3 (1.18) Not Important	3 (1.32) Not Important	0.84 (0.15)
Allergic Reaction to Contrast Dye - Moderate	4.6 (0.54) Important	4.4 (2.14) Important	4.4 (1.61) Important	4.1 (1.93) Important	5.8 (1.66) Important	4.9 (1.68) Important	3.7 (1.98) Important	4.8 (1.38) Important	4.8 (1.42) Important	4.6 (1.91) Important	4.3 (1.2) Important	0.71 (0.18)
Allergic Reaction to Contrast Dye - Severe	5.9 (0.35) Important	5.8 (2.17) Important	5.8 (2.08) Important	5.7 (2.39) Important	6.5 (1.57) Critical	6.1 (2.27) Important	5.1 (2.17) Important	6.2 (1.56) Important	5.9 (1.14) Important	6.2 (2.36) Important	5.8 (1.64) Important	0.53 (0.24)
Any Diagnostic Test - False Negative	5.9 (0.68) Important	5.6 (1.8) Important	5.9 (2.35) Important	5.3 (2.29) Important	5.2 (1.62) Important	6.3 (1.91) Important	4.6 (1.68) Important	6.3 (1.53) Important	5.9 (1.64) Important	6.6 (1.28) Critical	7 (1.94) Critical	0.6 (0.24)
Any Diagnostic Test - False Positive	5.7 (0.51) Important	5.4 (2.1) Important	5.9 (2.15) Important	5.6 (2.25) Important	5 (1.15) Important	5.7 (1.42) Important	4.9 (1.55) Important	5.8 (1.47) Important	5.5 (1.74) Important	6.3 (2) Important	6.6 (1.58) Critical	0.62 (0.21)

Any Diagnostic Test - Inconclusive	5 (0.61) Important	5.3 (2.49) Important	4.3 (2.34) Important	5.4 (2.19) Important	5 (1.83) Important	4 (1.28) Important	4.6 (1.59) Important	5.1 (1.45) Important	5 (1.84) Important	5.3 (1.55) Important	6.3 (1.83) Important	0.69 (0.18)
Any Diagnostic Test - True Negative	5.9 (0.56) Important	5.9 (2.1) Important	4.9 (3.03) Important	5.6 (2.23) Important	5.6 (1.95) Important	6.2 (2.44) Important	5.9 (1.73) Important	6.7 (1.7) Critical	5.2 (2.66) Important	6.3 (2.24) Important	6.8 (1.56) Critical	0.91 (0.15)
Any Diagnostic Test - True Positive	6.3 (0.78) Important	6 (2.52) Important	5.2 (2.85) Important	6.1 (2.22) Important	5.6 (1.95) Important	7.4 (2.01) Critical	5.7 (1.83) Important	6.9 (1.6) Critical	5.8 (2.58) Important	6.8 (2.23) Critical	7.6 (1.58) Critical	0.76 (0.19)
Aortic Aneurysm	6.4 (0.52) Important	6 (2.09) Important	5.9 (2.15) Important	5.7 (1.97) Important	6.9 (1.14) Critical	6.9 (1.44) Critical	6.1 (1.55) Important	6.1 (2.29) Important	6 (2.6) Important	7.3 (1.42) Critical	6.8 (1.48) Critical	0.46 (0.23)
Bleeding during Oocyte Retrieval	3.8 (0.49) Important	3.6 (1.8) Important	3.1 (1.51) Not Important	3.6 (1.97) Important	3.9 (1.85) Important	3.3 (1.3) Not Important	4.6 (2.42) Important	3.7 (1.71) Important	3.2 (1.76) Not Important	4.4 (1.5) Important	4.3 (1.39) Important	0.81 (0.16)
Burden of Therapy - Injections in Pregnancy	4.4 (0.35) Important	4.3 (1.49) Important	4 (2) Important	4 (2.09) Important	4.8 (2.09) Important	4.3 (0.43) Important	4.6 (1.84) Important	4.6 (1.44) Important	3.9 (1.9) Important	5 (2.19) Important	4.6 (1.22) Important	0.73 (0.18)
Burden of Therapy – Diagnostic Procedure	4.2 (0.43) Important	3.8 (2.13) Important	4.4 (2.29) Important	4.8 (1.59) Important	4.8 (1.54) Important	3.6 (1.55) Important	4.1 (1.36) Important	4.5 (1.86) Important	3.7 (1.77) Important	3.7 (1.62) Important	4.4 (1.8) Important	0.79 (0.18)
Central Venous Line Dysfunction in an Infant	5.5 (0.39) Important	5.8 (1.46) Important	4.6 (2.2) Important	5.2 (1.7) Important	6 (1.41) Important	5.8 (1.2) Important	5.4 (1.85) Important	5.3 (1.35) Important	5.2 (1.19) Important	5.8 (1.89) Important	5.5 (1.12) Important	0.6 (0.22)
Cerebral Venous Thrombosis – Mild	7.1 (0.47) Critical	7.1 (0.64) Critical	6.8 (1.54) Critical	5.9 (1.66) Important	7.5 (1.02) Critical	7.2 (1.34) Critical	7.3 (1.48) Critical	6.9 (1.39) Critical	7.7 (0.91) Critical	7.4 (1.28) Critical	6.9 (1.54) Critical	0.47 (0.18)
Cerebral Venous Thrombosis – Severe	8.3 (0.41) Critical	8.6 (0.48) Critical	8.1 (1.14) Critical	7.5 (1.67) Critical	8.7 (0.46) Critical	7.8 (1.4) Critical	8.4 (0.73) Critical	8.1 (1.59) Critical	8.8 (0.36) Critical	8.6 (0.66) Critical	8.3 (0.66) Critical	0.24 (0.15)
Cesarean Wound Complication	4.8 (0.44) Important	4.9 (1.71) Important	4.3 (1.42) Important	5.3 (1.29) Important	5.2 (1.69) Important	5 (0.71) Important	4 (1.6) Important	5.3 (1.48) Important	4.4 (1.82) Important	4.4 (1.28) Important	4.9 (1.69) Important	0.71 (0.18)
Chronic Kidney Disease	6.2 (0.47) Important	7 (0.91) Critical	6.3 (2) Important	6 (1.22) Important	6.5 (1.43) Critical	5.5 (1.37) Important	5.4 (1.68) Important	6.6 (1.14) Critical	6 (1.3) Important	6.3 (2.49) Important	6.6 (0.99) Critical	0.4 (0.17)
Chronic Liver Failure	6.5 (0.51) Critical	7.3 (1.11) Critical	6.7 (1.76) Critical	5.8 (1.64) Important	7.2 (1.47) Critical	6 (1.73) Important	6.3 (1.98) Important	6.6 (1.2) Critical	5.9 (2.02) Important	6.8 (2.36) Critical	6.8 (0.97) Critical	0.31 (0.17)
Chronic Pain - Mild	4.5 (0.54) Important	3.9 (1.55) Important	4.1 (2.27) Important	4.5 (1.12) Important	5 (1.84) Important	4.6 (1.07) Important	5.7 (1.75) Important	4.6 (1.54) Important	4.8 (1.56) Important	4 (1.61) Important	4 (1.58) Important	0.68 (0.16)
Chronic Pain - Moderate	6.2 (0.5) Important	6.1 (1.75) Important	5.3 (2.3) Important	6.3 (1.01) Important	7.1 (1.37) Critical	6.5 (1.08) Important	6.9 (2.17) Critical	6.3 (1.44) Important	6.2 (1.31) Important	6.2 (1.89) Important	5.6 (1.22) Important	0.45 (0.19)
Chronic Thrombotic Pulmonary Hypertension	7.4 (0.39) Critical	7.6 (1.04) Critical	7.1 (1.93) Critical	7.2 (1.52) Critical	7.5 (1.2) Critical	7.5 (0.99) Critical	7.6 (0.9) Critical	6.8 (1.22) Critical	7.2 (0.89) Critical	8.3 (1) Critical	7.1 (1.69) Critical	0.34 (0.15)
Cognitive Impairment - Mild	5.8 (0.56) Important	5.8 (1.86) Important	5.3 (2.26) Important	4.9 (1.66) Important	5.9 (1.45) Important	6.8 (1.34) Critical	5.6 (1.84) Important	5.5 (1.41) Important	6 (2.39) Important	6.6 (2.24) Critical	5.4 (1.8) Important	0.47 (0.21)
Cognitive Impairment - Severe	7.3 (0.54) Critical	7.5 (1.85) Critical	6.5 (2.39) Important	6.8 (1.36) Critical	7.6 (1.8) Critical	8.3 (1.35) Critical	7.6 (1.18) Critical	6.7 (1.81) Critical	7.6 (1.9) Critical	7.8 (2.32) Critical	7.3 (1.71) Critical	0.22 (0.18)
Congenital Malformation	7.2 (0.43) Critical	7.7 (1.7) Critical	6.8 (2.44) Critical	7.5 (1.16) Critical	8.1 (1.45) Critical	7 (2.6) Critical	7 (1.67) Critical	7 (2.24) Critical	6.9 (2.37) Critical	7 (2.32) Critical	6.6 (2.29) Critical	0.28 (0.18)
Contrast Induced Nephropathy	5.5 (0.32) Important	5.7 (1.43) Important	5.5 (1.5) Important	5.2 (2.37) Important	6.1 (1.58) Important	5.3 (2.18) Important	5.7 (1.03) Important	5.4 (1.54) Important	4.8 (1.56) Important	5.5 (1.91) Important	5.6 (0.99) Important	0.56 (0.2)
Critical INR	6 (0.43) Important	5.7 (1.31) Important	5.5 (1.83) Important	5.8 (2.13) Important	6.7 (1.35) Critical	6.2 (1.99) Important	6.4 (1.92) Important	6.5 (1.78) Important	6.2 (1.62) Important	5.4 (2.33) Important	5.6 (1.49) Important	0.68 (0.2)

CVC-Related Thrombosis in Adolescents - Severe	6.7 (0.46) Critical	6.9 (0.76) Critical	6.1 (1.51) Important	6.4 (2.14) Important	6.7 (1.62) Critical	6.4 (1.22) Important	6.6 (1.36) Critical	6.3 (1.55) Important	7.2 (0.8) Critical	7.7 (1.27) Critical	7 (1.22) Critical	0.48 (0.19)
CVC-Related Thrombosis in Infants	6.3 (0.46) Important	6.6 (0.95) Critical	5.8 (1.85) Important	5.9 (2.07) Important	6.6 (1.62) Critical	5.6 (0.86) Important	7 (1.1) Critical	5.8 (1.86) Important	6.8 (1.31) Critical	6.7 (1.73) Critical	6.5 (1.8) Critical	0.49 (0.22)
Deep Venous Thrombosis (DVT) in the Forearm – Mild	4.7 (0.75) Important	4.5 (1.04) Important	3.6 (1.6) Important	3.6 (2.18) Important	6.2 (1.4) Important	4.5 (1.72) Important	4.9 (2.1) Important	4.9 (2.05) Important	5.5 (1.65) Important	4.9 (2.47) Important	4.5 (2.24) Important	0.78 (0.15)
Deep Venous Thrombosis (DVT) in the Forearm – Moderate	5.6 (0.66) Important	5.5 (0.87) Important	5.1 (1.77) Important	4.8 (1.92) Important	7.1 (1.66) Critical	5.4 (1.07) Important	5.6 (2.06) Important	5.7 (1.81) Important	6.4 (1.39) Important	5.9 (2.12) Important	5 (2.18) Important	0.68 (0.18)
Deep Venous Thrombosis (DVT) in the Forearm – Severe	6.4 (0.59) Important	6.4 (1.11) Important	5.8 (1.87) Important	5.7 (1.49) Important	7.5 (1.57) Critical	6.3 (0.86) Important	6.3 (1.98) Important	6.6 (1.74) Critical	7.3 (1.26) Critical	6.8 (1.54) Critical	5.8 (1.98) Important	0.56 (0.2)
Deep Venous Thrombosis (DVT) in the Lower Leg – Mild	5.3 (0.71) Important	4.9 (1.19) Important	4.7 (1.49) Important	3.9 (2.14) Important	6.6 (1.5) Critical	5.2 (1.64) Important	5.4 (1.59) Important	5.7 (1.44) Important	6 (1.75) Important	5.7 (1.95) Important	5.1 (2.03) Important	0.77 (0.15)
Deep Venous Thrombosis (DVT) in the Lower Leg – Moderate	6.4 (0.56) Important	6.1 (0.95) Important	6.1 (1.44) Important	5.3 (1.83) Important	7.5 (1.2) Critical	6.3 (1.54) Important	6.4 (1.18) Important	6.7 (1) Critical	6.7 (1.43) Critical	6.8 (1.47) Critical	6.1 (1.9) Important	0.64 (0.16)
Deep Venous Thrombosis (DVT) in the Lower Leg – Severe	7.2 (0.5) Critical	7 (0.71) Critical	7.2 (1.17) Critical	6.2 (1.77) Important	8.1 (1.14) Critical	7.1 (1.16) Critical	7.4 (1.18) Critical	7.5 (0.91) Critical	7.5 (1.04) Critical	7.6 (1.07) Critical	6.7 (1.48) Critical	0.52 (0.17)
Deep Venous Thrombosis (DVT) in the Upper Arm – Mild	5.6 (0.61) Important	5.4 (1.19) Important	5 (1.36) Important	4.3 (2.05) Important	6.6 (1.43) Critical	5.4 (1.67) Important	5.7 (1.75) Important	5.7 (1.66) Important	6.3 (1.26) Important	5.9 (2.02) Important	5.5 (2) Important	0.75 (0.15)
Deep Venous Thrombosis (DVT) in the Upper Arm – Moderate	6.6 (0.42) Critical	6.6 (0.86) Critical	6.4 (0.84) Important	5.8 (1.48) Important	7.3 (1.27) Critical	6.5 (1.23) Critical	6.4 (1.76) Important	6.7 (1.19) Critical	7.2 (0.86) Critical	6.8 (1.6) Critical	6.3 (1.56) Important	0.61 (0.16)
Deep Venous Thrombosis (DVT) in the Upper Arm – Severe	7.5 (0.35) Critical	7.5 (0.65) Critical	7.5 (0.84) Critical	6.8 (0.9) Critical	7.9 (1.14) Critical	7.5 (0.78) Critical	7.4 (1.59) Critical	7.3 (1.19) Critical	8.2 (0.58) Critical	7.6 (1.02) Critical	7.4 (1.22) Critical	0.48 (0.17)
Deep Venous Thrombosis (DVT) in the Upper Leg – Mild	6.1 (0.5) Important	5.9 (1.38) Important	5.7 (1.73) Important	5.2 (1.75) Important	7 (1.1) Critical	6.2 (1.8) Important	5.7 (1.67) Important	6.6 (0.88) Critical	6.3 (1.14) Important	6.6 (1.56) Critical	6.1 (1.76) Important	0.71 (0.17)
Deep Venous Thrombosis (DVT) in the Upper Leg – Moderate	7.1 (0.41) Critical	7 (1) Critical	7.2 (1.41) Critical	6.5 (1.19) Critical	7.8 (0.87) Critical	6.8 (1.34) Critical	6.6 (1.59) Critical	7.5 (0.62) Critical	7.5 (1.01) Critical	7.6 (1.28) Critical	7.1 (1.36) Critical	0.58 (0.14)
Deep Venous Thrombosis (DVT) in the Upper Leg – Severe	8 (0.27) Critical	8.3 (0.43) Critical	8 (0.96) Critical	7.8 (0.83) Critical	8.4 (0.66) Critical	7.7 (1.21) Critical	7.9 (0.83) Critical	8.2 (0.65) Critical	8.4 (0.74) Critical	8.2 (0.75) Critical	7.6 (1.22) Critical	0.43 (0.16)
Delay of Intervention	4.6 (0.47) Important	4.9 (1.66) Important	4.3 (1.49) Important	4.7 (2.21) Important	5.3 (2.05) Important	5.2 (2.08) Important	3.7 (1.75) Important	4.9 (1.91) Important	4.3 (1.79) Important	4.2 (1.99) Important	4.8 (1.71) Important	0.78 (0.17)

Elevated Liver Values	4.3 (0.29) Important	4.3 (1.36) Important	4.5 (2.23) Important	4 (2.52) Important	4.8 (2.04) Important	4.4 (1.23) Important	4 (0.93) Important	4.8 (1.82) Important	4 (1.41) Important	4.1 (1.97) Important	4.1 (1.17) Important	0.78 (0.15)
Emergency Room Visit	4.6 (0.7) Important	4.7 (2.05) Important	4.6 (1.56) Important	4.2 (1.86) Important	6 (1.73) Important	4.1 (1.68) Important	4.3 (1.83) Important	4.9 (1.75) Important	4.5 (1.34) Important	3.3 (2) Not Important	5.4 (1.73) Important	0.75 (0.18)
Fetal Radiation Exposure	4.9 (0.66) Important	4.7 (2.29) Important	4.1 (2.26) Important	5.7 (2.26) Important	4.4 (1.57) Important	4.3 (1.48) Important	4.6 (1.02) Important	5.6 (1.73) Important	4.6 (2.17) Important	4.5 (2.29) Important	6.1 (1.54) Important	0.69 (0.23)
Gastrointestinal Bleeding - Major	7.4 (0.42) Critical	7.4 (1.38) Critical	7.3 (1.1) Critical	7.7 (1.18) Critical	8.1 (1.04) Critical	7.4 (1.15) Critical	7 (1.07) Critical	7.9 (0.68) Critical	6.8 (0.97) Critical	7.8 (0.87) Critical	7 (0.71) Critical	0.44 (0.19)
Gastrointestinal Bleeding - Minor	5.1 (0.53) Important	5.3 (1.42) Important	4.9 (1.97) Important	4.5 (1.62) Important	6.4 (1.43) Important	5.3 (0.86) Important	5 (1.6) Important	5.1 (1.75) Important	4.5 (1.28) Important	5.5 (1.63) Important	5 (1.32) Important	0.71 (0.16)
Heart or Lung Transplantation	7.1 (0.53) Critical	7.4 (2.06) Critical	7.2 (2.48) Critical	6.2 (2.62) Important	7.1 (2.07) Critical	7.5 (1.66) Critical	6.4 (1.96) Important	6.9 (1.89) Critical	6.5 (2.37) Critical	7.6 (1.91) Critical	7.9 (1.54) Critical	0.26 (0.16)
Hemiplegia	7.8 (0.54) Critical	7.8 (2.23) Critical	8.4 (0.64) Critical	7.2 (0.69) Critical	8.5 (0.81) Critical	8 (1.13) Critical	8 (0.93) Critical	7.7 (1.58) Critical	6.7 (2.05) Critical	8.2 (1.4) Critical	7.4 (1.41) Critical	0.21 (0.14)
Hemorrhagic Stroke	8.5 (0.23) Critical	8.6 (0.64) Critical	8.2 (1.27) Critical	8.3 (0.92) Critical	8.7 (0.64) Critical	8.6 (0.64) Critical	8.6 (0.49) Critical	8.3 (0.93) Critical	8.2 (1.46) Critical	8.5 (0.92) Critical	8.9 (0.33) Critical	0.12 (0.1)
Heparin Skin Necrosis	5.7 (0.53) Important	5.2 (1.67) Important	4.9 (1.73) Important	4.9 (1.75) Important	6.2 (2.32) Important	5.8 (1.64) Important	6 (1.51) Important	6.5 (1.54) Important	5.5 (1.78) Important	6.2 (2.32) Important	6 (1.22) Important	0.56 (0.19)
Heparin-Associated Bone Fractures due to Osteoporosis	5.5 (0.62) Important	5.9 (1.04) Important	5.2 (1.66) Important	5.2 (1.99) Important	6.6 (1.11) Critical	5.4 (1.77) Important	5 (1.31) Important	4.7 (1.81) Important	6.2 (1.58) Important	5.8 (1.47) Important	4.6 (1.65) Important	0.53 (0.19)
Heparin-Induced Thrombocytopenia (HIT)	6.6 (0.79) Critical	7.5 (1.04) Critical	6.3 (1.49) Important	5.8 (1.95) Important	7.7 (1.19) Critical	7.8 (1.03) Critical	6 (1.31) Important	6.9 (1.18) Critical	5.5 (2.27) Important	6.5 (1.8) Critical	6.1 (1.62) Important	0.49 (0.2)
HIT Test – False Negative	6.6 (0.75) Critical	7.6 (0.64) Critical	5.5 (2.39) Important	6.3 (1.81) Important	6.4 (1.36) Important	7.7 (0.86) Critical	5.6 (1.2) Important	6.6 (1.27) Critical	6.4 (1.82) Important	7.5 (1.12) Critical	6.3 (2.33) Important	0.49 (0.23)
HIT Test – False Positive	5.7 (0.47) Important	6.5 (1.71) Critical	5.1 (2.35) Important	5.3 (1.66) Important	5.3 (1.68) Important	5.9 (1.44) Important	5.6 (1.74) Important	6 (1.47) Important	5.7 (1.86) Important	6.4 (1.43) Important	5.3 (1.56) Important	0.68 (0.2)
Hospitalization Adults/Adolescents	5.1 (0.78) Important	5.4 (2.29) Important	5.1 (1.76) Important	5 (1.96) Important	6.6 (2.01) Critical	3.9 (1.38) Important	4.4 (1.4) Important	6.1 (1.15) Important	5.1 (0.73) Important	4.3 (1.85) Important	4.8 (1.3) Important	0.71 (0.21)
Hospitalization Children	5.3 (0.54) Important	4.8 (2.19) Important	4.8 (1.6) Important	5.2 (1.85) Important	6.4 (1.83) Important	4.8 (0.66) Important	5.4 (1.62) Important	5.6 (1.55) Important	4.6 (1.55) Important	5.7 (2.05) Important	5.6 (1.65) Important	0.57 (0.22)
Inadequate Patient Medication Adherence	5.6 (0.39) Important	6 (1.91) Important	5.5 (1.99) Important	5.1 (2.36) Important	5.6 (1.85) Important	5.5 (1.97) Important	5 (1.6) Important	6.1 (1.71) Important	5.6 (1.44) Important	5.5 (1.75) Important	6.3 (1.09) Important	0.76 (0.17)
Increased Duration of Hospitalization	5.1 (0.76) Important	5.8 (1.57) Important	5.2 (1.4) Important	4.9 (1.93) Important	6.7 (1.49) Critical	4.5 (1.37) Important	4 (1.6) Important	5.5 (1.41) Important	4.8 (1.34) Important	4.2 (1.33) Important	5 (1.94) Important	0.7 (0.2)
Infant Bleeding - Mild	4.9 (0.61) Important	4.5 (1.5) Important	4.6 (1.96) Important	5.2 (1.85) Important	6.3 (1.94) Important	4.8 (1.2) Important	4.4 (2.15) Important	4.5 (1.56) Important	4.2 (1.35) Important	5.4 (1.36) Important	5.3 (1.71) Important	0.67 (0.21)
Infant Bleeding - Severe	8.1 (0.75) Critical	8.2 (0.69) Critical	8.4 (0.68) Critical	8.3 (0.75) Critical	8.7 (0.67) Critical	8.1 (0.6) Critical	6 (2.9) Important	7.9 (1.98) Critical	8.1 (0.92) Critical	8.9 (0.3) Critical	8.3 (0.83) Critical	0.26 (0.19)
Interacting Medications	5.4 (0.62) Important	5.6 (1.71) Important	4.8 (2.62) Important	4.7 (2.01) Important	5.7 (2.1) Important	5.6 (1.87) Important	4.7 (1.75) Important	6.9 (1.41) Critical	5.4 (2.02) Important	5 (2.14) Important	5.5 (1.5) Important	0.74 (0.17)

Intracardiac Shunt (Glenn or Fontan) Thrombosis in a Child	7.4 (0.72) Critical	7.8 (1.27) Critical	7.3 (2.05) Critical	6.6 (2.06) Critical	7.7 (1.56) Critical	7.5 (0.87) Critical	6 (2.53) Important	6.6 (1.92) Critical	8.2 (0.86) Critical	8.4 (0.92) Critical	7.9 (1.05) Critical	0.27 (0.14)
Ischemic Stroke - Mild	7 (0.61) Critical	6.3 (1.89) Important	6.3 (1.35) Important	6.4 (1.5) Important	7.3 (1.19) Critical	7.6 (0.98) Critical	8 (0.93) Critical	6.5 (1.5) Critical	6.8 (0.97) Critical	7.8 (0.75) Critical	7 (1.41) Critical	0.39 (0.19)
Ischemic Stroke - Severe	8.3 (0.39) Critical	7.8 (2.07) Critical	8.2 (1.03) Critical	7.6 (1.93) Critical	8.6 (0.66) Critical	8.6 (0.48) Critical	8.9 (0.35) Critical	8.1 (1.31) Critical	8.2 (1.12) Critical	8.8 (0.4) Critical	8.4 (0.7) Critical	0.14 (0.1)
IVC Filter Failure	6.2 (0.37) Important	6.3 (1.42) Important	5.8 (1.19) Important	5.7 (1.43) Important	6.1 (1.97) Important	6.4 (1.61) Important	6.8 (1.95) Critical	6.1 (1.14) Important	5.8 (1.19) Important	6.5 (1.75) Critical	6.6 (1.11) Critical	0.56 (0.18)
Limb Amputation	7.7 (0.6) Critical	8.2 (1.67) Critical	7.7 (1.96) Critical	6.5 (1.71) Critical	8 (1.1) Critical	8.2 (1.19) Critical	7.6 (0.73) Critical	8.1 (1.55) Critical	6.6 (2.56) Critical	8.2 (1.54) Critical	7.5 (1.5) Critical	0.26 (0.16)
Low Time in Therapeutic INR Range; High INR Variability	5.4 (0.52) Important	5 (1.78) Important	5 (2.13) Important	5.3 (2.05) Important	6.3 (1.27) Important	5.1 (2.07) Important	5.7 (1.75) Important	6.1 (1.57) Important	5.9 (1.59) Important	4.7 (2) Important	5 (1) Important	0.74 (0.18)
Major Bleeding	8.4 (0.3) Critical	8.7 (0.47) Critical	8.2 (0.58) Critical	8.3 (0.75) Critical	8.7 (0.64) Critical	8.5 (0.5) Critical	7.7 (0.7) Critical	8.5 (0.62) Critical	8.2 (0.8) Critical	8.8 (0.4) Critical	8.3 (0.97) Critical	0.33 (0.23)
Maternal Breast Irradiation	4.2 (0.55) Important	4.6 (2.02) Important	3.3 (2) Not Important	4.2 (2.41) Important	4 (1.49) Important	3.5 (1.41) Important	3.8 (1.33) Important	4.8 (1.72) Important	4.1 (1.54) Important	4.5 (2.01) Important	5.1 (0.93) Important	0.79 (0.17)
Maternal Plasma Drug Anticoagulation Level	3.7 (0.66) Important	2.9 (1.75) Not Important	2.8 (1.9) Not Important	3.7 (2.3) Important	4.4 (2.11) Important	3.4 (1.41) Not Important	4.8 (1.47) Important	4.5 (1.71) Important	3.1 (1.98) Not Important	3.8 (2.23) Important	3.6 (1.58) Important	0.87 (0.14)
Maternal Skin Reactions	3.5 (0.66) Important	3.2 (1.77) Not Important	2.5 (1.5) Not Important	4 (1.54) Important	4.5 (2.16) Important	2.9 (0.78) Not Important	4.4 (2.33) Important	4.1 (1.38) Important	3.2 (1.46) Not Important	3.4 (1.43) Not Important	3 (1.5) Not Important	0.8 (0.16)
Mesenteric Vein Thrombosis – Acute	7.4 (0.57) Critical	7.7 (1.11) Critical	7.4 (1.5) Critical	6.8 (1.64) Critical	7.9 (1.04) Critical	7.4 (1.07) Critical	7.3 (1.58) Critical	6.7 (1.78) Critical	6.9 (1.14) Critical	8.7 (0.46) Critical	7.1 (1.45) Critical	0.38 (0.17)
Mesenteric Vein Thrombosis – Sub-Acute	6.1 (0.5) Important	6.2 (1.67) Important	6.1 (1.7) Important	5.3 (1.37) Important	7 (1) Critical	6.5 (1.5) Important	6 (1.51) Important	5.7 (1.88) Important	6 (0.96) Important	6.8 (1.6) Critical	5.5 (1.58) Important	0.53 (0.16)
Mild Anxiety	3.5 (0.58) Important	3.5 (1.44) Important	2.7 (1.71) Not Important	3.4 (1.85) Not Important	4.5 (1.96) Important	2.7 (0.96) Not Important	4.3 (2.05) Important	4.1 (1.65) Important	3.7 (1.2) Important	3.1 (1.58) Not Important	3.4 (1.11) Not Important	0.84 (0.11)
Minor Bleeding	4.5 (0.68) Important	4.2 (1.57) Important	4.4 (1.44) Important	4.4 (1.61) Important	5.7 (1.42) Important	4.8 (1.75) Important	3.1 (1.88) Not Important	5.3 (1.53) Important	3.8 (1.31) Important	4.5 (1.63) Important	4.6 (1.22) Important	0.81 (0.15)
Moderate to Severe Anxiety	5 (0.61) Important	5.3 (1.93) Important	3.7 (1.71) Important	4.5 (1.88) Important	5.7 (1.55) Important	4.4 (1.37) Important	5.9 (1.73) Important	5.1 (1.86) Important	5.2 (1.17) Important	5.1 (1.51) Important	4.9 (1.45) Important	0.65 (0.18)
Multiple Organ Failure	7.8 (0.37) Critical	7.8 (1.62) Critical	7.7 (1.96) Critical	6.9 (2.47) Critical	8.1 (1.14) Critical	7.9 (1.93) Critical	7.3 (2.21) Critical	7.7 (1.29) Critical	7.8 (1.79) Critical	8 (1.9) Critical	8.3 (0.66) Critical	0.15 (0.14)
Neonatal Bleeding - Mild	4.8 (0.72) Important	4.8 (1.88) Important	4.1 (1.58) Important	5.5 (1.97) Important	6.3 (1.94) Important	4.1 (1.36) Important	4.2 (2.4) Important	4.6 (1.49) Important	4 (1.52) Important	5 (1.1) Important	5.4 (1.49) Important	0.65 (0.23)
Neonatal Bleeding - Severe	7.9 (0.43) Critical	8.3 (0.72) Critical	7.8 (0.87) Critical	8.2 (0.94) Critical	8.6 (0.68) Critical	7.8 (1.09) Critical	7 (2.1) Critical	7.7 (2.05) Critical	7.5 (1.28) Critical	8.3 (0.64) Critical	8 (0.87) Critical	0.3 (0.21)

Osteoporosis	4.8 (0.42) Important	4.3 (1.69) Important	4.8 (1.59) Important	4.4 (1.55) Important	5.2 (1.4) Important	4.5 (1.67) Important	4.9 (1.25) Important	4.5 (1.89) Important	5.7 (1.32) Important	5.2 (2.09) Important	4.5 (0.87) Important	0.68 (0.18)
Peripheral Arterial Disease	6.3 (0.65) Important	5.8 (2.31) Important	6.2 (1.19) Important	5 (1.68) Important	6.8 (1.4) Critical	7.2 (1.03) Critical	6 (1.41) Important	6.7 (1.69) Critical	6.1 (2.2) Important	7.2 (1.25) Critical	6.3 (1.2) Important	0.45 (0.16)
Permanent Disability or Dependency	7.7 (0.41) Critical	8.1 (0.67) Critical	8.1 (0.79) Critical	7.3 (0.72) Critical	8.3 (0.9) Critical	7.7 (0.86) Critical	7.4 (1.92) Critical	7.5 (1.45) Critical	7.3 (1.49) Critical	8.3 (0.9) Critical	7.3 (1.09) Critical	0.23 (0.16)
Placental Abruption – Non-Severe	5.1 (0.46) Important	5.6 (0.95) Important	4.3 (1.66) Important	5.2 (1.64) Important	5.9 (2.18) Important	4.6 (1.49) Important	5 (2) Important	5.3 (1.49) Important	4.7 (1.81) Important	5.4 (2.5) Important	5.4 (1.32) Important	0.69 (0.19)
Placental Abruption – Severe	6.8 (0.46) Critical	6.8 (1.34) Critical	6.3 (2.26) Important	7.1 (1.38) Critical	6.8 (1.81) Critical	6.9 (1.9) Critical	6.4 (1.99) Important	7.2 (1.86) Critical	5.8 (2.32) Important	7 (2.28) Critical	7.5 (1) Critical	0.48 (0.24)
Portal Vein Thrombosis – Acute	6.6 (0.56) Critical	6.7 (2.09) Critical	6.2 (1.54) Important	6.4 (1.8) Important	7.3 (1.35) Critical	6.6 (1.37) Critical	6.9 (1.25) Critical	6.2 (1.64) Important	5.7 (1.81) Important	7.8 (1.25) Critical	6.6 (1.58) Critical	0.52 (0.19)
Portal Vein Thrombosis – Chronic	5.8 (0.7) Important	6 (1.73) Important	5.2 (2.18) Important	5.7 (1.37) Important	7.3 (1.42) Critical	5.6 (1.55) Important	5.7 (2.05) Important	5 (1.71) Important	5.5 (1.74) Important	6.6 (2.06) Critical	4.9 (1.27) Important	0.59 (0.17)
Portal Vein Thrombosis in a Child	7.1 (0.52) Critical	7.3 (0.75) Critical	7 (1.95) Critical	6.1 (2.07) Important	7.3 (1.35) Critical	7.5 (1.12) Critical	7.2 (1.83) Critical	6.3 (1.7) Important	7.8 (1.36) Critical	7.7 (1) Critical	7.3 (0.97) Critical	0.4 (0.18)
Post-Thrombotic Syndrome	6.4 (0.36) Important	5.8 (2.09) Important	6.1 (1.88) Important	6.1 (1.26) Important	7 (1.26) Critical	6.4 (1.23) Important	6.9 (1.55) Critical	6.2 (1.42) Important	6.3 (1.07) Important	6.6 (1.43) Critical	6.3 (0.66) Important	0.61 (0.17)
Preeclampsia	5.7 (0.55) Important	6.3 (1.3) Important	4.7 (1.42) Important	6.1 (1.78) Important	5.7 (1.33) Important	4.9 (1.17) Important	6.3 (1.28) Important	6.1 (1.66) Important	5.2 (2.01) Important	5.8 (2.14) Important	6.1 (1.36) Important	0.62 (0.19)
Pregnancy Loss	6.5 (0.82) Important	6.1 (2.66) Important	7.2 (2.17) Critical	7.2 (0.94) Critical	8 (1.41) Critical	5.3 (2.44) Important	7 (1.51) Critical	6.8 (2.17) Critical	6.1 (2.46) Important	5.5 (2.73) Important	5.9 (2.03) Important	0.4 (0.25)
Preterm Delivery	5.8 (0.46) Important	6.3 (1.89) Important	4.9 (1.98) Important	6 (1.91) Important	6 (2) Important	6 (2.24) Important	5.2 (1.72) Important	5.8 (1.53) Important	5.3 (1.83) Important	6.3 (2) Important	5.9 (1.36) Important	0.63 (0.21)
Preterm Labor	4.8 (0.44) Important	4.3 (2.29) Important	4.2 (1.7) Important	5.4 (1.77) Important	4.8 (2.2) Important	5.3 (1.71) Important	4.8 (1.47) Important	5.3 (1.6) Important	4.2 (1.7) Important	5.1 (1.7) Important	4.8 (1.3) Important	0.73 (0.18)
Prolonged Immobilization	6 (0.62) Important	6.7 (1.03) Critical	5.8 (0.94) Important	5.8 (1.4) Important	6.9 (1.58) Critical	6.4 (0.64) Important	4.6 (1.5) Important	5.9 (1.41) Important	5.6 (1.27) Important	6 (1.67) Important	6.4 (1.58) Important	0.53 (0.2)
Psychological Burden of Diagnostic Labels	4.6 (0.75) Important	4.8 (1.77) Important	3.5 (1.37) Not Important	4.3 (2.24) Important	5 (1.1) Important	4 (1.13) Important	6.3 (1.98) Important	4.7 (1.58) Important	3.8 (1.61) Important	4.5 (1.96) Important	5 (1.73) Important	0.74 (0.17)
Pulmonary Embolism - Mild	6.7 (0.33) Critical	6.5 (1.32) Critical	6.2 (1.67) Important	6.3 (1.88) Important	7.2 (1.03) Critical	6.5 (1.62) Critical	6.6 (1.92) Critical	7.1 (0.77) Critical	6.7 (1.2) Critical	7.1 (1.37) Critical	6.6 (1.8) Critical	0.62 (0.16)
Pulmonary Embolism - Moderate	7.9 (0.14) Critical	7.9 (0.64) Critical	7.8 (1.12) Critical	7.9 (0.86) Critical	8.2 (0.6) Critical	7.9 (0.9) Critical	7.7 (1.28) Critical	7.9 (0.72) Critical	7.9 (0.62) Critical	8.1 (0.94) Critical	8 (0.71) Critical	0.42 (0.15)
Pulmonary Embolism - Severe	8.8 (0.12) Critical	8.8 (0.43) Critical	8.8 (0.58) Critical	8.9 (0.28) Critical	8.8 (0.4) Critical	8.8 (0.4) Critical	8.4 (0.73) Critical	8.8 (0.4) Critical	8.8 (0.36) Critical	8.7 (0.46) Critical	8.8 (0.43) Critical	0.25 (0.14)
Pulmonary Function	5.6 (0.83) Important	5.9 (2.06) Important	3.8 (1.78) Important	5.3 (2.13) Important	6.2 (1.83) Important	5.6 (2.19) Important	6.6 (1.84) Critical	6.1 (1.71) Important	5.8 (1.1) Important	4.5 (1.63) Important	6.4 (1.11) Important	0.65 (0.23)
Quality of Life Impairment	6.8 (0.52) Critical	7.3 (0.94) Critical	5.9 (1.62) Important	6.7 (1.31) Critical	7.3 (1.1) Critical	6.3 (1.54) Important	7.6 (1.4) Critical	7.3 (1.24) Critical	6.8 (1.1) Critical	6.3 (1.79) Important	6.6 (1.11) Critical	0.57 (0.2)

Radiation Exposure	3.9 (0.52) Important	4.1 (1.85) Important	3.4 (2.06) Not Important	4.2 (2.19) Important	4.2 (1.17) Important	2.9 (1) Not Important	4.1 (1.46) Important	3.9 (1.34) Important	4.7 (1.64) Important	3.4 (1.28) Not Important	4.5 (1.22) Important	0.84 (0.14)
Renal Vein Thrombosis in a Child - Bilateral	7.7 (0.59) Critical	7.8 (0.94) Critical	7.3 (2) Critical	6.9 (1.98) Critical	7.6 (1.5) Critical	8.3 (0.97) Critical	7.2 (1.33) Critical	6.7 (1.93) Critical	8.3 (1.07) Critical	8.5 (0.67) Critical	8 (0.71) Critical	0.28 (0.17)
Renal Vein Thrombosis in a Child - Unilateral	6.6 (0.43) Critical	7 (0.85) Critical	6.2 (1.89) Important	6.2 (2.04) Important	7.2 (1.25) Critical	7.1 (1.17) Critical	6.2 (1.6) Important	6.1 (1.78) Important	6.5 (1.34) Critical	7.1 (1.64) Critical	6.8 (1.09) Critical	0.5 (0.2)
Reoperation	6.4 (0.64) Important	6.3 (1.25) Important	7.6 (1.15) Critical	5.6 (1.55) Important	6.8 (1.6) Critical	7.1 (1.17) Critical	5.6 (2.06) Important	6.4 (1.73) Important	6 (1.24) Important	6 (1.61) Important	7 (0.71) Critical	0.56 (0.2)
Retinal Vein Occlusion	6.2 (0.43) Important	5.7 (1.89) Important	6.3 (1.91) Important	5.6 (2.19) Important	6.5 (2.16) Critical	5.9 (1.45) Important	6.6 (1.84) Critical	5.6 (1.85) Important	6.3 (1.54) Important	6.5 (1.96) Critical	6.9 (1.05) Critical	0.49 (0.19)
Skin Complications from Compression Stockings	4.3 (0.39) Important	4 (1.68) Important	4 (1.71) Important	4.3 (1.92) Important	4.5 (2.06) Important	3.8 (1.11) Important	3.9 (2.03) Important	4.5 (1.67) Important	5 (2.08) Important	4.9 (2.39) Important	4.4 (0.99) Important	0.77 (0.14)
Small for Gestational Age	5.2 (0.38) Important	5.6 (1.66) Important	4.5 (2.02) Important	5.6 (1.49) Important	5.6 (1.85) Important	4.9 (2.09) Important	4.7 (1.28) Important	5.6 (1.66) Important	5.2 (1.87) Important	5 (2.19) Important	5.3 (1.2) Important	0.65 (0.2)
Spinal Epidural Hematoma	7.4 (0.53) Critical	7.3 (2.09) Critical	7.5 (1.62) Critical	7.5 (1.23) Critical	8.1 (1.51) Critical	7.5 (1.66) Critical	6.4 (1.62) Important	6.9 (2.1) Critical	7.4 (1.6) Critical	8.4 (1.02) Critical	7.3 (1.39) Critical	0.37 (0.22)
Splenomegaly	4.4 (0.47) Important	4.8 (1.62) Important	4.2 (1.64) Important	3.7 (1.84) Important	5.1 (1.87) Important	4.2 (2.08) Important	3.7 (2.25) Important	4.6 (1.96) Important	4.7 (1.86) Important	4.1 (2.47) Important	4.9 (1.05) Important	0.72 (0.16)
Systemic-to-Pulmonary Shunt Thrombosis in an Infant	7.4 (0.73) Critical	7.7 (1.49) Critical	7.2 (1.99) Critical	6.9 (2.11) Critical	7.8 (1.62) Critical	7.5 (1.58) Critical	5.8 (2.14) Important	6.6 (1.92) Critical	8.5 (0.63) Critical	7.9 (1.3) Critical	7.8 (1.39) Critical	0.25 (0.16)
Venous Ulcer	6.3 (0.37) Important	6 (2.16) Important	5.8 (1.9) Important	6.5 (1.04) Critical	6.5 (1.5) Critical	6.7 (1.29) Critical	5.7 (1.28) Important	6.5 (1.2) Important	6.2 (1.72) Important	6.9 (1.76) Critical	6.5 (1.12) Critical	0.53 (0.19)
Vulvar Hematoma	4.6 (0.78) Important	4.8 (1.59) Important	3.3 (1.6) Not Important	4.5 (1.62) Important	4.6 (1.77) Important	5.5 (1.5) Important	3.9 (2.03) Important	4.5 (1.66) Important	4 (1.66) Important	6.2 (1.33) Important	4.5 (1.12) Important	0.71 (0.19)
Wound Hematoma	5 (0.56) Important	5.1 (1.66) Important	5.8 (1.56) Important	4.5 (1.43) Important	6 (1.73) Important	5.4 (1.93) Important	4.2 (1.72) Important	4.8 (1.74) Important	4.5 (1.28) Important	4.8 (1.81) Important	4.9 (1.17) Important	0.73 (0.17)
Wound Infection	5.3 (0.54) Important	5.5 (1.61) Important	6.3 (1.98) Important	5.6 (1.67) Important	6.2 (1.78) Important	5.4 (1.22) Important	4.8 (2.14) Important	4.8 (1.62) Important	4.8 (1.29) Important	4.8 (1.54) Important	5.3 (1.2) Important	0.67 (0.17)
Acne - Mild	-	-	-	-	-	-	-	-	-	-	-	0.89 (0.13)
Acne - Severe	-	-	-	-	-	-	-	-	-	-	-	0.75 (0.19)
Dysmenorrhea	-	-	-	-	-	-	-	-	-	-	-	0.73 (0.18)
Irregular Menses	-	-	-	-	-	-	-	-	-	-	-	0.8 (0.17)
Menorrhagia	-	-	-	-	-	-	-	-	-	-	-	0.73 (0.18)
Unintended Pregnancy	-	-	-	-	-	-	-	-	-	-	-	0.61 (0.24)
Mean	6.01	6.04	5.68	5.68	6.56	6.02	5.83	6.10	5.89	6.23	6.07	0.56

Standard Deviation	1.25	1.38	1.47	1.21	1.23	1.46	1.31	1.16	1.36	1.51	1.28	0.19
Minimum	3.19	2.92	2.45	3.00	3.89	2.73	2.86	3.47	3.08	3.00	3.00	0.12
Maximum	8.76	8.75	8.77	8.92	8.80	8.80	8.86	8.80	8.85	8.90	8.88	0.91

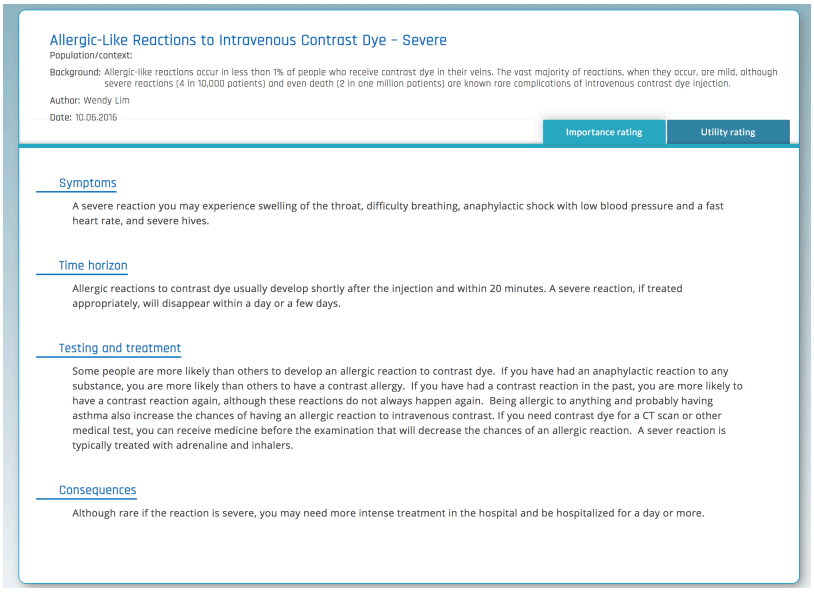
Table 3: Comparison of panels' outcome utility ratings to those reported in the literature

Outcome or Health Outcome Descriptor	Panelists' utility rating Mean (SD)	Utility ranges reported in the literature*
Pulmonary embolism		0.63-0.93
Pulmonary embolism - severe	0.25 (0.14)	
Pulmonary embolism - moderate	0.42 (0.15)	
Pulmonary embolism - mild	0.62 (0.16)	
Deep vein thrombosis		0.61-0.99
Deep vein thrombosis in the upper leg – severe	0.43(0.16)	
Deep vein thrombosis in the upper leg – moderate	0.58 (0.14)	
Deep vein thrombosis in the upper leg – mild	0.71 (0.17)	
Minor Bleeding	0.81 (0.15)	
Major Bleeding	0.32 (0.23)	
Major intracranial bleeding		0.15
Central nervous system bleeding		0.29-0.60
Muscular bleeding		0.76
Minor intracranial bleeding		0.75
Gastrointestinal tract bleeding	0.44 (0.19)	0.59-0.65
Post-thrombotic syndrome	0.61 (0.17)	0.82
Post-thrombotic syndrome - severe		0.93-0.98
Post-thrombotic syndrome - mild		0.99-1.00

* utility value ranges across studies reported in systematic review of patients' values and preferences (Etxeandia-Ikobaltzeta *et al.* 2020)[13]; a value of 1 represents the health state of 'full health' and a value of 0 represents the health state of being 'dead'.

Chapter 3: Figures

Figure 1. Example of health outcome descriptors



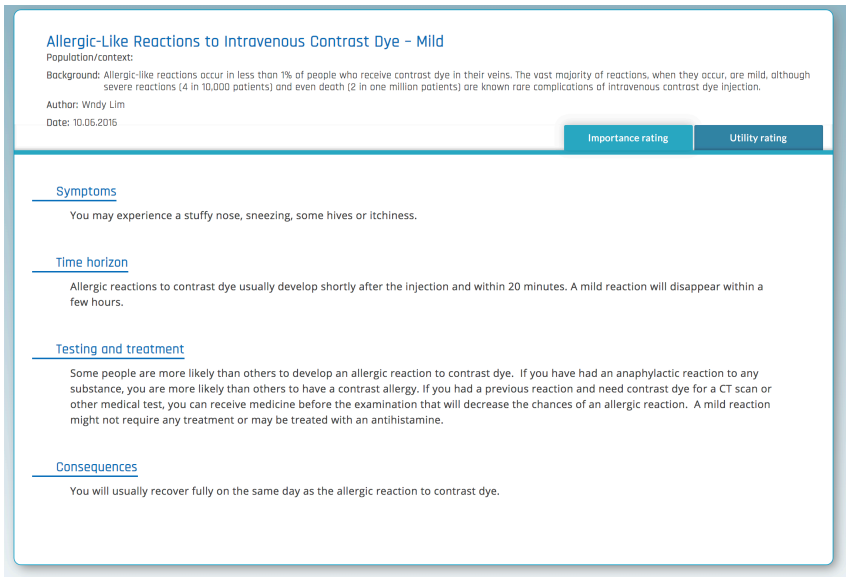


Figure 1 Caption: provided a definition with respect to symptoms, time horizon, testing and treatment, and consequences for a person who experiences the health outcome. (from ms.gradepro.org)

Figure 2. Visual analogue scale and instructions for panel members to rate the utility of health outcomes

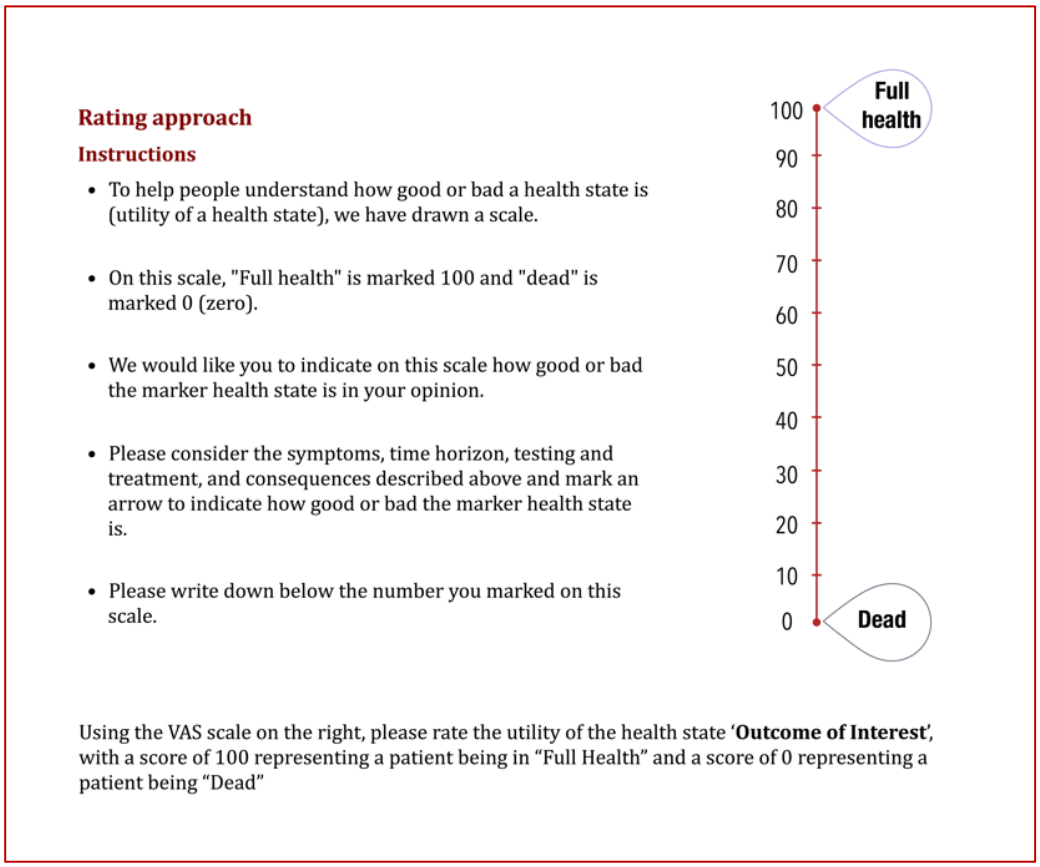


Figure 3. Question importance rating categories across the 10 guideline panels

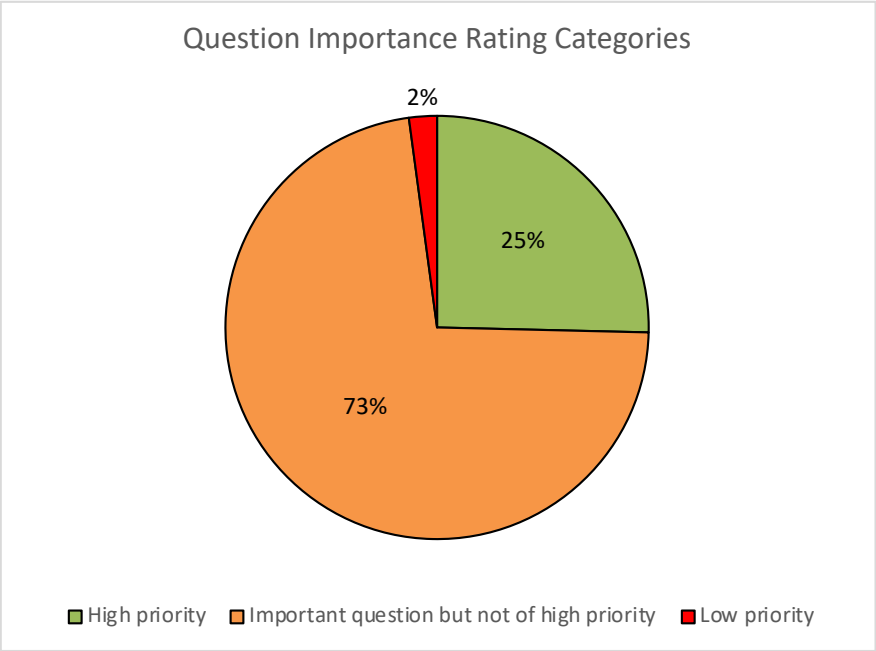


Figure 3 Caption: This figure shows the distribution of question importance categories based on panels' mean ratings. Rating of 1 to 3 – Low priority; Rating of 4 to 6 – Important question but not of high priority; Rating of 7 to 9 – High priority.

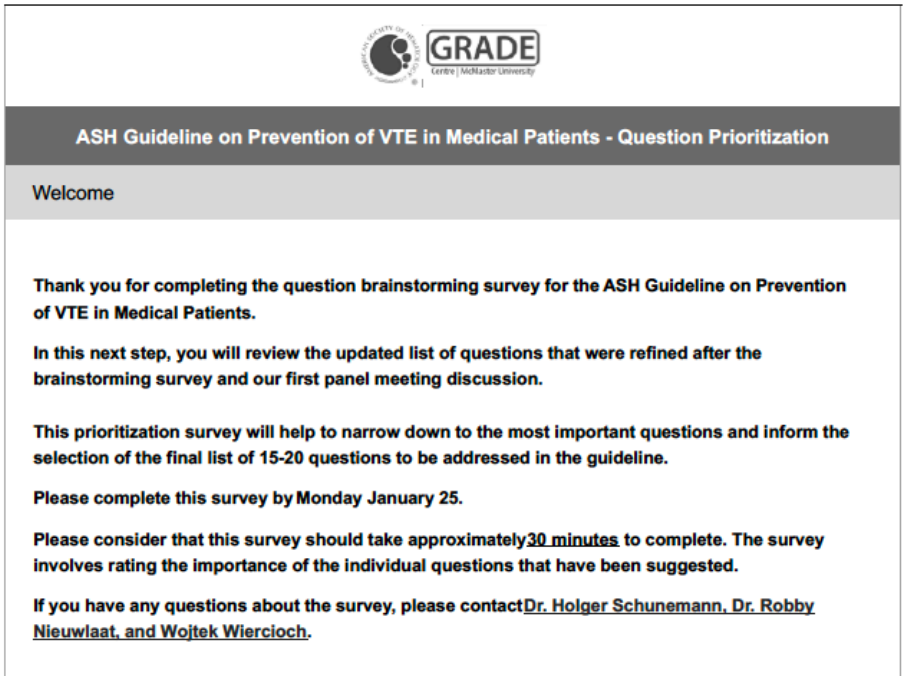
Chapter 3: Appendix A: Guideline Topics and Question Prioritization

Table 1: Guideline Topics and Question Prioritization in the ASH VTE Guidelines

Guideline Topic	Number of Panel Members	Number of Questions Rated	Minutes to Complete Survey*
			Mean (SD)
Prevention of VTE in Medical Patients	12	35	37 (22)
Diagnosis of VTE	12	19	37 (21)
Optimal Management of Anticoagulation	13	52	38 (18)
Heparin-Induced Thrombocytopenia	12	45	44 (24)
VTE in the Context of Pregnancy	11	33	47 (22)
Treatment of Pediatric VTE	16	112	62 (24)
Treatment of VTE	15	75	56 (31)
VTE in Patients with Cancer	16	20	38 (30)
Prevention of VTE in Surgical Patients	15	41	28 (11)
Thrombophilia Testing	9	37	28 (14)

*The mean time to complete the survey excludes 13 panel members with a completion time of greater than 3 hours who presumably took a break and returned to the survey.

Chapter 3: Appendix B: Question Prioritization Rating Survey Example



GRADE Centre | McMaster University

ASH Guideline on Prevention of VTE in Medical Patients - Question Prioritization

Welcome

Thank you for completing the question brainstorming survey for the ASH Guideline on Prevention of VTE in Medical Patients.

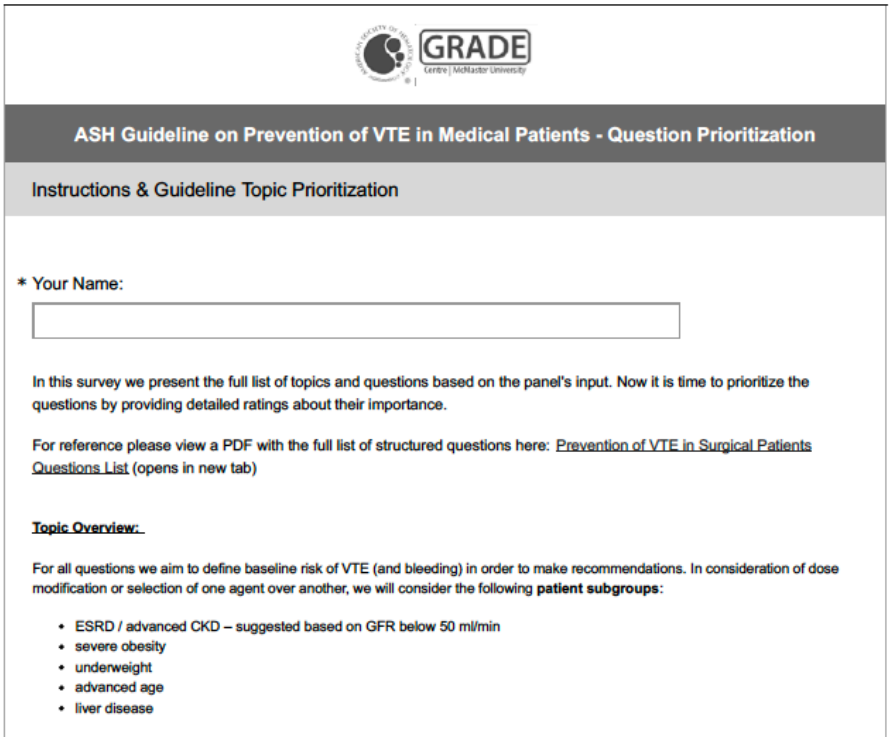
In this next step, you will review the updated list of questions that were refined after the brainstorming survey and our first panel meeting discussion.

This prioritization survey will help to narrow down to the most important questions and inform the selection of the final list of 15-20 questions to be addressed in the guideline.

Please complete this survey by Monday January 25.

Please consider that this survey should take approximately 30 minutes to complete. The survey involves rating the importance of the individual questions that have been suggested.

If you have any questions about the survey, please contact Dr. Holger Schunemann, Dr. Robby Nieuwlaat, and Wojtek Wiercioch.



GRADE Centre | McMaster University

ASH Guideline on Prevention of VTE in Medical Patients - Question Prioritization

Instructions & Guideline Topic Prioritization

* Your Name:

In this survey we present the full list of topics and questions based on the panel's input. Now it is time to prioritize the questions by providing detailed ratings about their importance.

For reference please view a PDF with the full list of structured questions here: Prevention of VTE in Surgical Patients Questions List (opens in new tab)

Topic Overview:

For all questions we aim to define baseline risk of VTE (and bleeding) in order to make recommendations. In consideration of dose modification or selection of one agent over another, we will consider the following **patient subgroups**:

- ESRD / advanced CKD – suggested based on GFR below 50 ml/min
- severe obesity
- underweight
- advanced age
- liver disease

For the populations and interventions of interest, we will consider the following:

Identifying risk: How do we best identify medical inpatients at:

1. Higher VTE risk
2. Higher bleeding risk
3. PICC or Central Line specifically

Populations - Inpatient:

1. Acutely ill
2. Stroke
3. Critically ill
4. Chronically ill (chronic vent, etc.)
5. Dicharged post medical services to home or rehab
6. Cancer (to decide whether to split out or address in context of risk scores)
7. Patients hospitalized with active major bleeding

Populations - Outpatient:

1. People with a previous thrombosis who are not on long-term prevention treatment and who have transient risk factors.
2. Ambulatory medical patients with other risk factors such as advance age, nursing home, minor immobility/injuries, outpatient

2

illness/infection
3. Long distance travelers >4 hours
4. Asymptomatic relatives or others with inherited thrombophilic disorders
5. Cancer (to be discussed with the cancer guideline)
6. Endurance athletes

Interventions – Mechanical:

- Graduated compression stockings (GCS)
- Intermittent pneumatic compression (IPC)
- Venous foot pump
- IVC filter

Interventions - Pharmacological:

- VKA, LMWH, fondaparinux, LDUH, ASA, DOACs, statins

Prioritizing the list of topics:

In the first part of the survey below please rate the importance of the overall topics/populations to help prioritize them. Using the 9-point Likert type scale presented below, select a rating from 1 (least important) to 9 (most important). Please consider the relative importance of the topics to each other (i.e. selecting a rating of 8 or 9 for all topics will not provide informative ratings for prioritization).

Prioritize Guideline Topic Sections: Rate the importance of each of the proposed sections/populations for the VTE Prevention in Medical Patients guideline

Inpatient:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
I. Acutely ill patients	○	○	○	○	○	○	○	○	○
II. Stroke patients	○	○	○	○	○	○	○	○	○
III. Critically ill patients	○	○	○	○	○	○	○	○	○
IV. Chronically ill patients	○	○	○	○	○	○	○	○	○
V. Discharged post medical services to home or rehab	○	○	○	○	○	○	○	○	○
VI. Cancer	○	○	○	○	○	○	○	○	○
VII. Patients hospitalized with active major bleeding	○	○	○	○	○	○	○	○	○
VIII. General Questions (DOACs, VKA, ASA, statins)	○	○	○	○	○	○	○	○	○

3

Outpatient:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
I. People with a previous thrombosis who are not on long-term prevention treatment and who have transient risk factors.	○	○	○	○	○	○	○	○	○
II. Ambulatory medical patients with other risk factors such as advance age, nursing home, minor immobility/injuries, outpatient illness/infection	○	○	○	○	○	○	○	○	○
III. Long distance travelers >4 hours	○	○	○	○	○	○	○	○	○
IV. Asymptomatic relatives or others with inherited thrombophilic disorders	○	○	○	○	○	○	○	○	○
V. Cancer	○	○	○	○	○	○	○	○	○
VI. Endurance athletes	○	○	○	○	○	○	○	○	○

Do you have any comments about prioritizing the overall guideline topics?

Click Next to proceed to the question prioritization.

Question Prioritization - Question 1 Sub-Questions and Overall Importance

Section 1: Acutely Ill Medical Patients

For this question, using the 9-point scale first rate the importance of the question according to the 6 criteria outlined. Select a rating for each criterion to judge the question as being of least or most importance with respect to the question being one.. :

i. That commonly arises in practice,
ii. For which there is uncertainty in practice regarding how to manage patients,
iii. For which there is new research evidence to consider,
iv. That is associated with variation in practice,
v. That has important consequences for, or is associated with, high resource use or costs,
vi. That has not been previously or sufficiently addressed (e.g. in guidelines)

Then, select an overall importance rating. Please also note that the 9-point scales are not a ranking, so the same number can be assigned more than once.

Q1. What absolute risk of symptomatic VTE should be considered high risk in acutely ill medical patients?

How important is this question with respect to the following criteria?:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9	Don't know	N/A
1. Common question in practice?	○	○	○	○	○	○	○	○	○	○	○
2. Uncertainty in practice?	○	○	○	○	○	○	○	○	○	○	○
3. New evidence to consider?	○	○	○	○	○	○	○	○	○	○	○
4. Variation in practice?	○	○	○	○	○	○	○	○	○	○	○
5. Consequences for resource use/cost?	○	○	○	○	○	○	○	○	○	○	○
6. Not previously or sufficiently addressed?	○	○	○	○	○	○	○	○	○	○	○

Select an overall importance rating for the question:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Click Next to proceed to rating the next set of questions.

Question Prioritization for the Guideline

Prioritizing the list of questions - continued:

1. To assist with determining the importance of each question below, using the 9-point Likert-type scale please rate each question according to the 6 criteria outlined. Select a rating for each criterion to judge the question as being of least or most importance with respect to the question being one.. :

   i. That commonly arises in practice,
   ii. For which there is uncertainty in practice regarding how to manage patients,
   iii. For which there is new research evidence to consider,
   iv. That is associated with variation in practice,
   v. That has important consequences for, or is associated with, high resource use or costs,
   vi. That has not been previously or sufficiently addressed (e.g. in guidelines)

2. Then, select an overall importance rating for the question from 1 (least important) to 9 (most important) using the 9 point Likert-type scale. For this rating, please consider the relative importance of **all** the questions presented (i.e. selecting a rating of 8 or 9 for all questions will not provide informative ratings for prioritization). Please also note that the 9-point scales are not a ranking, so the same number can be assigned more than once.

Use the space provided to add any comments. Your judgments and comments will be used to reduce the questions list and inform the selection of the final questions to be addressed in the guideline.

Section 1: Acutely Ill Medical Patients - Continued

In acutely ill medical patients...:

Q2. What absolute risk of hemorrhage should be considered a contraindication to pharmacologic prophylaxis?:

How important is this question with respect to the following criteria?:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9	Don't know	N/A
1. Common question in practice?	○	○	○	○	○	○	○	○	○	○	○
2. Uncertainty in practice?	○	○	○	○	○	○	○	○	○	○	○
3. New evidence to consider?	○	○	○	○	○	○	○	○	○	○	○
4. Variation in practice?	○	○	○	○	○	○	○	○	○	○	○
5. Consequences for resource use/cost?	○	○	○	○	○	○	○	○	○	○	○
6. Not previously or sufficiently addressed?	○	○	○	○	○	○	○	○	○	○	○

Select an overall importance rating for the question:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Q3. Should heparin versus no heparin be used for prophylaxis of VTE?:

How important is this question with respect to the following criteria?:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9	Don't know	N/A
1. Common question in practice?	○	○	○	○	○	○	○	○	○	○	○
2. Uncertainty in practice?	○	○	○	○	○	○	○	○	○	○	○
3. New evidence to consider?	○	○	○	○	○	○	○	○	○	○	○
4. Variation in practice?	○	○	○	○	○	○	○	○	○	○	○
5. Consequences for resource use/cost?	○	○	○	○	○	○	○	○	○	○	○
6. Not previously or sufficiently addressed?	○	○	○	○	○	○	○	○	○	○	○

Select an overall importance rating for the question:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Q4. Should LMWH versus UFH be used for prophylaxis of VTE?:

How important is this question with respect to the following criteria?:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9	Don't know	N/A
1. Common question in practice?	○	○	○	○	○	○	○	○	○	○	○
2. Uncertainty in practice?	○	○	○	○	○	○	○	○	○	○	○
3. New evidence to consider?	○	○	○	○	○	○	○	○	○	○	○
4. Variation in practice?	○	○	○	○	○	○	○	○	○	○	○
5. Consequences for resource use/cost?	○	○	○	○	○	○	○	○	○	○	○
6. Not previously or sufficiently addressed?	○	○	○	○	○	○	○	○	○	○	○

Select an overall importance rating for the question:

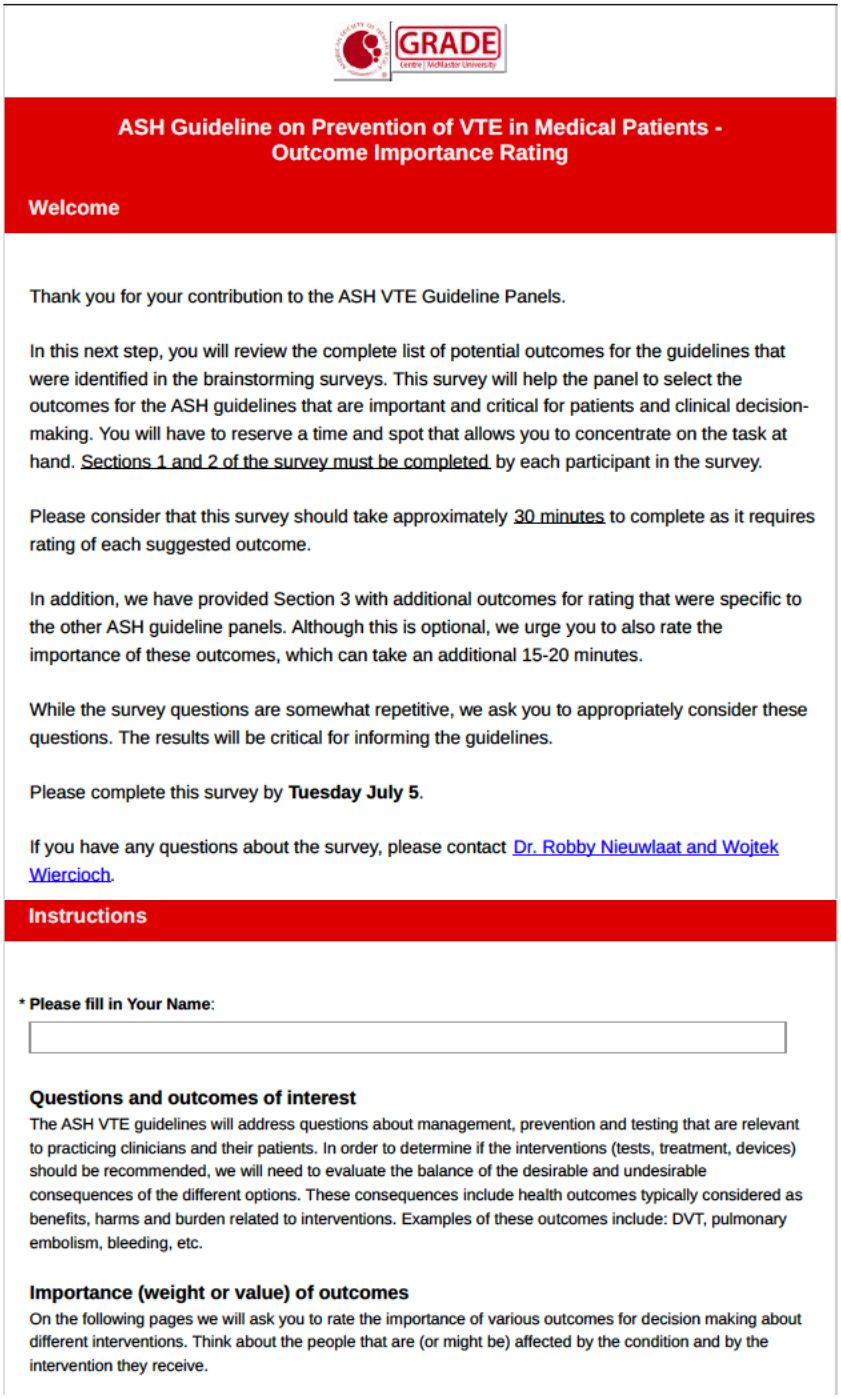
	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Chapter 3: Appendix C: Question Prioritization Survey Results Example

Legend for Colour Coding (mean rounded to nearest digit):	high priority to address in the guideline (rating of 7 to 9)			an important question but not of high priority (rating of 4 to 6)			question of low priority (rating of 1 to 3)					
Your Name:		**Panelist 1**	**Panelist 2**	**Panelist 3**	**Panelist 4**	**Panelist 5**	**Panelist 6**	**Panelist 7**	**Panelist 8**	**Panelist 9**	**Panelist 10**	**Mean Rating**	**Median Rating:**
Prioritize Guideline Topic Sections :	**I. Acutely ill patients**	8	8	9	9	9	9	8	9	8	9	**8.60**	**9**
	II. Stroke patients	6	8	8	9	6	9	7	8	8	6	**7.50**	**8**
	III. Critically ill patients	6	8	9	9	9	9	6	8	8	8	**8.00**	**8**
	IV. Chronically ill patients	5	8	5	4	9	5	7	8	5	2	**5.80**	**5**
	V. Discharged post medical services to hom	7	8	7	4	9	5	4	6	3	3	**5.60**	**5.5**
	VI. Cancer	5	8	8	7	6	9	6	8	8	5	**7.00**	**7.5**
	VII. Patients hospitalized with active major l	9	8	3	8	6	8	4	8	8	7	**6.90**	**8**
	VIII. General Questions (DOACs, VKA, ASA,	4	6	9	8	9	6	7	7	7	4	**6.70**	**7**
Outpatient:	**I. People with a previous thrombosis who ar**	8	8	8	8	9	8	6	8	7	9	**7.90**	**8**
	II. Ambulatory medical patients with other r	6	8	7	7	6	5	5	7	5	7	**6.30**	**6.5**
	III. Long distance travelers >4 hours	8	6	6	3	9	5	4	3	5	6	**5.50**	**5.5**
	IV. Asymptomatic relatives or others with ir	3	5	5	3	6	6	5	3	3	5	**4.40**	**5**
	V. Cancer	7	8	9	6	6	9	6	8	8	8	**7.50**	**8**
	VI. Endurance athletes	9	4	5	3	6	2	2	3	2	4	**4.00**	**3.5**
	Women on OCP (suggested addition)												
Do you have any comments about prioritizing the overall guideline topics?	**Open-Ended Response**	This is a very comprehensive list. For the outpatient po			I am not prior	Low enthusia	Priority should be given to populations which contribute to the overall national VTE burden. Further priority						
Q1. What absolute risk of symptomatic VTE should be considered high risk in acutely ill medical patients?	**1. Common question in practice?**		8	8	9	9	9	9	6	9	9	**8.44**	**9**
	2. Uncertainty in practice?		8	8	9		9	9	8	9	8	**8.50**	**8.5**
	3. New evidence to consider?		8	8	7	5	9	6	4		4	**6.38**	**6.5**
	4. Variation in practice?		8	8	8	9	9	9		8	7	**8.25**	**8**
	5. Consequences for resource use/cost?		6	6	8	9	9	9	8	9	5	**7.67**	**8**
	6. Not previously or sufficiently addressed?		7	7	6		9	9	6	9	6	**7.38**	**7**
Select an overall importance rating	**Overall importance**		7	8	9	9	9	9	8	8	8	**8.33**	**8**
Q2. What absolute risk of hemorrhage should be considered a contraindication to pharmacologic prophylaxis?:	**1. Common question in practice?**	9	8	8	8	9	9	9	8	8	9	**8.50**	**8.5**
	2. Uncertainty in practice?	8	8	8	9	7	9	9	8	8	8	**8.20**	**8**
	3. New evidence to consider?	7	6		7	3	7	4	4		7	**5.63**	**6.5**
	4. Variation in practice?		6	8	8	7	7	9	6		4	**6.88**	**7**
	5. Consequences for resource use/cost?	4	5	9	8	9	9	9	6	8	5	**7.20**	**8**
	6. Not previously or sufficiently addressed?		6	7	9		9	9	8	8	6	**7.75**	**8**
Select an overall importance rating	**Overall importance**		7	8	8	9	9	9	8	8	9	**8.33**	**8**
Acutely ill Medical Patients													
Q3. Should heparin versus no heparin be used for prophylaxis of VTE?:	**1. Common question in practice?**	9	8	7	8	9	9	6	6	1	7	**7.00**	**7.5**
	2. Uncertainty in practice?	8	7	7	7	7	9	6	6	1	6	**6.40**	**7**
	3. New evidence to consider?	8	6	7	4	3	9	6	2	1	4	**5.00**	**5**
	4. Variation in practice?		5	7	7	7	9	8	4	5	5	**6.33**	**7**
	5. Consequences for resource use/cost?	4	4	7	4	9	9	6	8	9	3	**6.30**	**6.5**
	6. Not previously or sufficiently addressed?		6	7	4	3	9	6	2	1	2	**4.44**	**4**
Select an overall importance rating	**Overall importance**	6	6	7		9	9	6	6	4		**6.63**	**6**

Q4. Should LMWH versus UFH be used for prophylaxis of VTE?:	1. Common question in practice?	8	6	1	7	6	9	4	6	6	9	6.20	6
	2. Uncertainty in practice?		6	2	4	3	9	4	6	6	8	5.33	6
	3. New evidence to consider?	9	5	1	4	3	9	3	2	2	6	4.40	3.5
	4. Variation in practice?		5	2	5	6	9	7	4	8	7	5.89	6
	5. Consequences for resource use/cost?	4	5	2	6	7	9	4	8	7	5	5.70	5.5
	6. Not previously or sufficiently addressed?		5	1	3	3	9	4	2	2	4	3.67	3
Select an overall importance rating	Overall importance	7	5	2	3	6	9	5	6	4	6	5.30	5.5
Q5. Should fondaparinux versus other pharmacologic agents be used for prophylaxis of VTE?:	1. Common question in practice?	8	6	1	3	6	6	7	5	1	4	4.70	5.5
	2. Uncertainty in practice?		6	2	3	3	6	8	5		6	4.88	5.5
	3. New evidence to consider?	9	5	1	3	3	4	4	2	1	5	3.70	3.5
	4. Variation in practice?		5	2	3	6	4	8	5		2	4.38	4.5
	5. Consequences for resource use/cost?	4	5	4	3	8	4	5	5	1	7	4.60	4.5
	6. Not previously or sufficiently addressed?		5	1	3	3	4	5	2	1	3	3.00	3
Select an overall importance rating	Overall importance	7	6	2	3	7	5	6	5	2	4	4.70	5
Q6. Should any DOAC versus other pharmacologic agents be used for prophylaxis of VTE?:	1. Common question in practice?	8	6	7		9	9	7	8	6	7	7.44	7
	2. Uncertainty in practice?		7	5		7	9	2	7	6	4	5.88	6.5
	3. New evidence to consider?	9	6	1		7	3	2	6	2	5	4.56	5
	4. Variation in practice?		6	1		7	3	7	6		3	4.71	6
	5. Consequences for resource use/cost?	4	6	5		9	8	5	8	9	8	6.89	8
	6. Not previously or sufficiently addressed?		6	7		7	3	2	6	6	6	5.38	6
Select an overall importance rating	Overall importance	8	6	5	1	9	3	4	7	6	7	5.60	6
Q7. Should extended duration (i.e., up to 30 or 40 days) versus a regular duration (i.e., in hospital only) be used for the thromboprophylaxis of VTE?:	1. Common question in practice?	7	7	7	7	9	9	7	8	5	8	7.40	7
	2. Uncertainty in practice?	8	7	8	7	8	9	8	7	5	7	7.40	7.5
	3. New evidence to consider?	9	6		3	6	7	6	5	5	6	5.89	6
	4. Variation in practice?		6	5	3	8	8	7	8	2	4	5.67	6
	5. Consequences for resource use/cost?	5	5	7	6	9	9	8	8	9	5	7.10	7.5
	6. Not previously or sufficiently addressed?		6	7	6	5	9	8	2	5	3	5.67	6
Select an overall importance rating	Overall importance	6	6	7	4	9	9	7	7	5	7	6.70	7
Q8. Should graduated compression stockings (GCS) versus no GCS be used for prophylaxis of VTE?	1. Common question in practice?	9	7	3	4	9	9	8	6	7	8	7.00	7.5
	2. Uncertainty in practice?	6	6	3	5	8	9	8	5	7	7	6.40	6.5
	3. New evidence to consider?	7	4		6	8	3	1	3	7	3	4.67	4
	4. Variation in practice?		5	2	5	8	9	8	6	7	6	6.22	6
	5. Consequences for resource use/cost?		5	2	4	8	9	2	8	7	4	5.44	5
	6. Not previously or sufficiently addressed?		6	2	4	7		2	3	7	5	4.50	4.5
Select an overall importance rating	Overall importance	4	5	2	2	9	9	4	6	7	5	5.30	5

Chapter 3: Appendix D: Outcome Importance Rating Survey Example



ASH Guideline on Prevention of VTE in Medical Patients - Outcome Importance Rating

Welcome

Thank you for your contribution to the ASH VTE Guideline Panels.

In this next step, you will review the complete list of potential outcomes for the guidelines that were identified in the brainstorming surveys. This survey will help the panel to select the outcomes for the ASH guidelines that are important and critical for patients and clinical decision-making. You will have to reserve a time and spot that allows you to concentrate on the task at hand. Sections 1 and 2 of the survey must be completed by each participant in the survey.

Please consider that this survey should take approximately 30 minutes to complete as it requires rating of each suggested outcome.

In addition, we have provided Section 3 with additional outcomes for rating that were specific to the other ASH guideline panels. Although this is optional, we urge you to also rate the importance of these outcomes, which can take an additional 15-20 minutes.

While the survey questions are somewhat repetitive, we ask you to appropriately consider these questions. The results will be critical for informing the guidelines.

Please complete this survey by **Tuesday July 5**.

If you have any questions about the survey, please contact Dr. Robby Nieuwlaat and Wojtek Wiercioch.

Instructions

*** Please fill in Your Name**:

Questions and outcomes of interest

The ASH VTE guidelines will address questions about management, prevention and testing that are relevant to practicing clinicians and their patients. In order to determine if the interventions (tests, treatment, devices) should be recommended, we will need to evaluate the balance of the desirable and undesirable consequences of the different options. These consequences include health outcomes typically considered as benefits, harms and burden related to interventions. Examples of these outcomes include: DVT, pulmonary embolism, bleeding, etc.

Importance (weight or value) of outcomes

On the following pages we will ask you to rate the importance of various outcomes for decision making about different interventions. Think about the people that are (or might be) affected by the condition and by the intervention they receive.

Suggestions and reminders

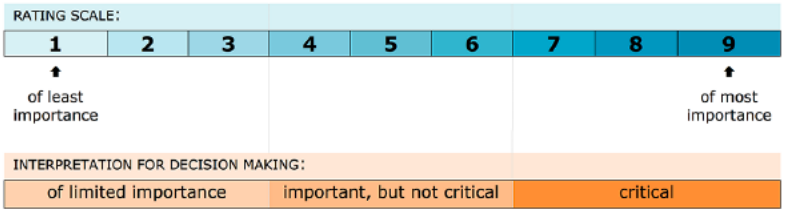
1. rate the importance of outcomes for decision-making in the context of the topic
2. rate the importance of outcomes irrespective of your belief in the effect of interventions on those outcomes
3. use your best knowledge on how important these outcomes would be for the people affected
4. you may rate outcomes as equally important (no need to rank order).

TASK

Rate the relative importance of the outcome for decision-making (i.e. formulating a recommendation) of each outcome on a scale from 1 to 9. The meaning of the ratings are:
1–3 are of limited or no importance for decision-making
4–6 are important, but not critical for decision-making
7–9 are critical for decision-making.

Again, you can use the same rating for different outcomes more than once.

Importance of an outcome



Click Next to proceed to the Outcome Importance Rating.

Section 1 - Outcomes Suggested by the Panel

Pulmonary Embolism - Severe

(Blood Clot in the Lung - Severe)

Symptoms: You will experience severe shortness of breath that will make it impossible for you to move and requires oxygen administration as well as tightness in your chest.
Time Horizon: Severe pulmonary embolism may impair you for the rest of your life, but can also resolve completely within weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will be administered in the hospital, frequently in a critical care unit. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. To identify the cause of the severe pulmonary embolism you may require additional testing such as blood work or other x-rays and similar tests.
Consequences: Consequences often include persisting shortness of breath, particularly with exercise. You are at immediate risk of dying with a severe pulmonary embolism.

Q1. Rate how important the outcome of **'Pulmonary Embolism - Severe'** is for making a decision about the optimal management strategy for diagnosing, preventing or treating Venous Thromboembolism:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Pulmonary Embolism - Moderate

(Blood Clot in the Lung - Moderate Severity)

Symptoms: You will experience shortness of breath, sometimes pain and tightness in your chest.
Time Horizon: Moderate pulmonary embolism will impair you for weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will be administered in the hospital for a few days or at home. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. You may require oxygen administration to improve your symptoms.To identify the cause of your problem you may require additional testing such as blood work or other x-rays and similar tests.
Consequences: You are at an increased risk of dying with a moderate pulmonary embolism. Consequences sometimes include persisting shortness of breath, particularly with exercise.

Q2. Rate how important the outcome of **'Pulmonary embolism - Moderate'** is for making a decision about the optimal management strategy for diagnosing, preventing or treating Venous Thromboembolism:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Pulmonary Embolism - Mild

(Blood Clot in the Lung - Mild Severity)

Symptoms: You may experience shortness of breath, sometimes pain and tightness in your chest develops if the problem is not immediately diagnosed.
Time Horizon: Mild pulmonary embolism will impair you for weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will typically be administered at home. It will include administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years.
Consequences: You are at a low risk of dying with a mild pulmonary embolism. The symptoms usually disappear.

Q3. Rate how important the outcome of **'Pulmonary embolism - Mild'** is for making a decision about the optimal management strategy for diagnosing, preventing or treating Venous Thromboembolism:

	Of LEAST importance 1	2	3	4	5	6	7	8	Of MOST importance 9
Overall importance	○	○	○	○	○	○	○	○	○

Deep Venous Thrombosis (DVT) in the Upper Leg – Severe

(Blood Clot in the Thigh - Severe)

Symptoms: You experience severe swelling, pain, warmth, heaviness or redness in your entire leg.
Time Horizon: Severe DVT will persist for months and will slowly improve.
Testing and Treatment: Treatment may be administered in the hospital or at home. Treatment typically includes administration of blood thinners using a small tube inserted in your vein, injections or pills. Long-lasting treatment with blood thinners is often required.
Consequences: Consequences often include long-lasting pain and swelling in the leg. Sometimes, it may also include a blood clot travelling to the lungs (a pulmonary embolism) and death.

Q4. Rate how important the outcome of **'DVT in the Upper Leg - Severe'** is for making a decision about the optimal management strategy for diagnosing, preventing or treating Venous Thromboembolism:

	Of LEAST importance								Of MOST importance
	1	2	3	4	5	6	7	8	9
Overall importance	○	○	○	○	○	○	○	○	○

Deep Venous Thrombosis (DVT) in the Upper Leg – Moderate

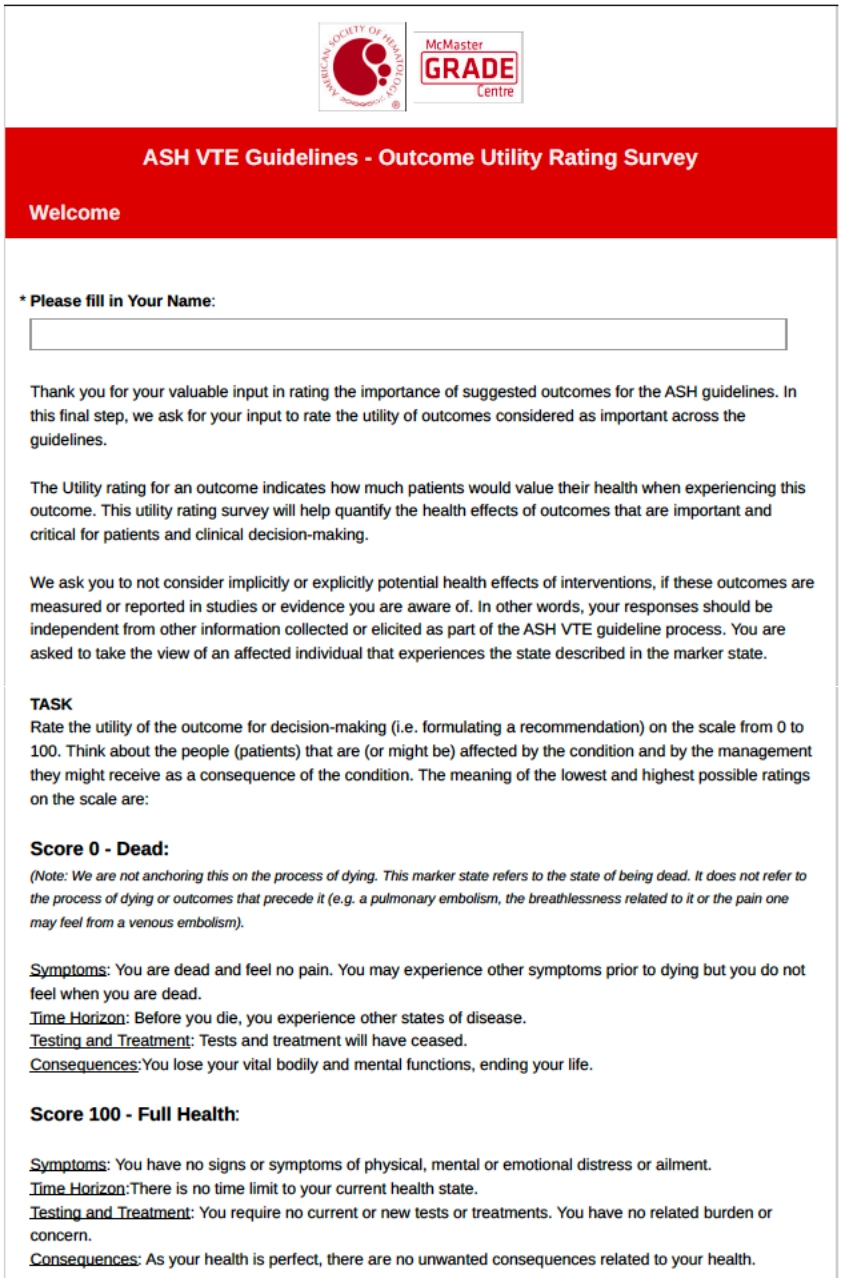
(Blood Clot in the Thigh - Moderate Severity)

Symptoms: You experience some swelling, pain, warmth, heaviness or redness in your entire leg.
Time Horizon: Moderate DVT will persist for months but improve over that time.
Testing and Treatment: Treatment may be administered in the hospital or at home. Treatment typically includes administration of blood thinners using a small tube inserted in your vein, injections or pills. Long-lasting treatment with blood thinners is often required.
Consequences: Consequences may include long-lasting pain and swelling in the leg. Rarely, it may also include a blood clot travelling to the lungs (a pulmonary embolism) and death.

Q5. Rate how important the outcome of **'DVT in the Upper Leg - Moderate'** is for making a decision about the optimal management strategy for diagnosing, preventing or treating Venous Thromboembolism:

	Of LEAST importance								Of MOST importance
	1	2	3	4	5	6	7	8	9
Overall importance	○	○	○	○	○	○	○	○	○

Chapter 3: Appendix E: Outcome Utility Rating Survey Example



ASH VTE Guidelines - Outcome Utility Rating Survey

Welcome

*** Please fill in Your Name**:

Thank you for your valuable input in rating the importance of suggested outcomes for the ASH guidelines. In this final step, we ask for your input to rate the utility of outcomes considered as important across the guidelines.

The Utility rating for an outcome indicates how much patients would value their health when experiencing this outcome. This utility rating survey will help quantify the health effects of outcomes that are important and critical for patients and clinical decision-making.

We ask you to not consider implicitly or explicitly potential health effects of interventions, if these outcomes are measured or reported in studies or evidence you are aware of. In other words, your responses should be independent from other information collected or elicited as part of the ASH VTE guideline process. You are asked to take the view of an affected individual that experiences the state described in the marker state.

TASK
Rate the utility of the outcome for decision-making (i.e. formulating a recommendation) on the scale from 0 to 100. Think about the people (patients) that are (or might be) affected by the condition and by the management they might receive as a consequence of the condition. The meaning of the lowest and highest possible ratings on the scale are:

Score 0 - Dead:
(Note: We are not anchoring this on the process of dying. This marker state refers to the state of being dead. It does not refer to the process of dying or outcomes that precede it (e.g. a pulmonary embolism, the breathlessness related to it or the pain one may feel from a venous embolism).

Symptoms: You are dead and feel no pain. You may experience other symptoms prior to dying but you do not feel when you are dead.
Time Horizon: Before you die, you experience other states of disease.
Testing and Treatment: Tests and treatment will have ceased.
Consequences:You lose your vital bodily and mental functions, ending your life.

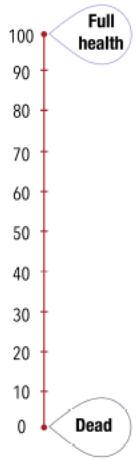
Score 100 - Full Health:

Symptoms: You have no signs or symptoms of physical, mental or emotional distress or ailment.
Time Horizon:There is no time limit to your current health state.
Testing and Treatment: You require no current or new tests or treatments. You have no related burden or concern.
Consequences: As your health is perfect, there are no unwanted consequences related to your health.

Rating approach

Instructions

- To help people understand how good or bad a health state is (utility of a health state), we have drawn a scale.
- On this scale, "Full health" is marked 100 and "dead" is marked 0 (zero).
- We would like you to indicate on this scale how good or bad the marker health state is in your opinion.
- Please consider the symptoms, time horizon, testing and treatment, and consequences described above and mark an arrow to indicate how good or bad the marker health state is.
- Please write down below the number you marked on this scale.



Using the VAS scale on the right, please rate the utility of the health state '**Outcome of Interest**', with a score of 100 representing a patient being in "Full Health" and a score of 0 representing a patient being "Dead"

Please Note:

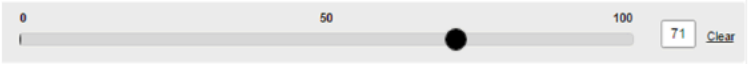
The VAS scale for you to select the utility will be depicted horizontally, but we ask you to keep the verrtical VAS scale in mind and follow the same instructions as provided above. We also acknowledge that the VAS is considered to not elicit utilities in the narrow definition, but we will use the term 'utility' for this survey. You can use the same value (utility) for different outcomes more than once. Please make use of the full scale and exact utility values for your selection (e.g. not only values of '10s' such as 10, 20, 40, 60, etc.). See here two examples of how you can rate the outcomes (examples taken from respiratory disease marker states):

EXAMPLE - Severe Lung Disease (Severe Chronic Obstructive Pulmonary Disease)

Please rate the utility of the health state **Severe Lung Disease**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".

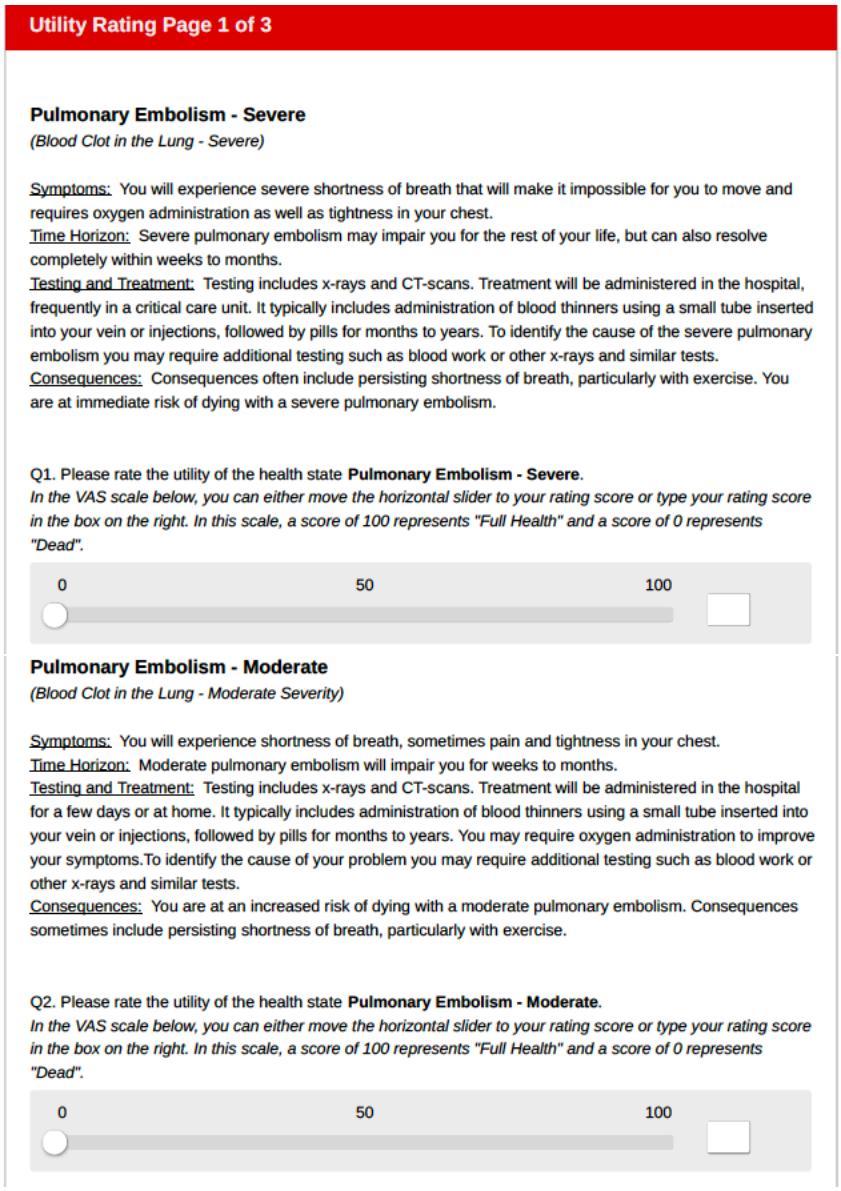
EXAMPLE – Mild Shortness of Breath (Mild Dyspnea)

Please rate the utility of the health state **Mild Shortness of Breath**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".



You will have to reserve a time and spot that allows you to concentrate on the task at hand. Please consider that this survey should take approximately 30 minutes to complete as it requires you to rate the utility of all the outcomes presented on 3 pages of the survey. You can complete the survey in multiple sittings by clicking 'Save and Go To Next' to save your progress, and return to the survey through your personal survey link in your email.

Click 'Save and Go To Next' to proceed to the Outcome Utility Rating.



Utility Rating Page 1 of 3

Pulmonary Embolism - Severe

(Blood Clot in the Lung - Severe)

Symptoms: You will experience severe shortness of breath that will make it impossible for you to move and requires oxygen administration as well as tightness in your chest.
Time Horizon: Severe pulmonary embolism may impair you for the rest of your life, but can also resolve completely within weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will be administered in the hospital, frequently in a critical care unit. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. To identify the cause of the severe pulmonary embolism you may require additional testing such as blood work or other x-rays and similar tests.
Consequences: Consequences often include persisting shortness of breath, particularly with exercise. You are at immediate risk of dying with a severe pulmonary embolism.

Q1. Please rate the utility of the health state **Pulmonary Embolism - Severe**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".

0 50 100

Pulmonary Embolism - Moderate

(Blood Clot in the Lung - Moderate Severity)

Symptoms: You will experience shortness of breath, sometimes pain and tightness in your chest.
Time Horizon: Moderate pulmonary embolism will impair you for weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will be administered in the hospital for a few days or at home. It typically includes administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years. You may require oxygen administration to improve your symptoms.To identify the cause of your problem you may require additional testing such as blood work or other x-rays and similar tests.
Consequences: You are at an increased risk of dying with a moderate pulmonary embolism. Consequences sometimes include persisting shortness of breath, particularly with exercise.

Q2. Please rate the utility of the health state **Pulmonary Embolism - Moderate**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".

0 50 100

Pulmonary Embolism - Mild
(Blood Clot in the Lung - Mild Severity)

Symptoms: You may experience shortness of breath, sometimes pain and tightness in your chest develops if the problem is not immediately diagnosed.
Time Horizon: Mild pulmonary embolism will impair you for weeks to months.
Testing and Treatment: Testing includes x-rays and CT-scans. Treatment will typically be administered at home. It will include administration of blood thinners using a small tube inserted into your vein or injections, followed by pills for months to years.
Consequences: You are at a low risk of dying with a mild pulmonary embolism. The symptoms usually disappear.

Q3. Please rate the utility of the health state **Pulmonary Embolism - Mild**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".

0 50 100

Deep Venous Thrombosis (DVT) in the Upper Leg – Severe
(Blood Clot in the Thigh - Severe)

Symptoms: You experience severe swelling, pain, warmth, heaviness or redness in your entire leg. Walking is very uncomfortable and/or painful for you.
Time Horizon: Severe DVT will persist for months and will slowly improve.
Testing and Treatment: You will require blood tests or radiological tests that are not painful. You will receive immediate initial treatment in hospital. Treatment typically includes administration of blood thinners using a small tube inserted in your vein, injections or pills. Long-lasting treatment with blood thinners is often required, possibly for the rest of your life.
Consequences: Consequences often include long-lasting pain and swelling in the leg. Sometimes, it may also include a blood clot travelling to the lungs (a pulmonary embolism) and death. You may be worried about side effects, including bleeding as a result of taking blood thinners. To reduce swelling you wear compression stockings most days. These stockings are difficult to put on and somewhat uncomfortable to wear but reduce swelling. You feel worried about having another leg clot periodically on most days for about 3 months and very seldom thereafter.

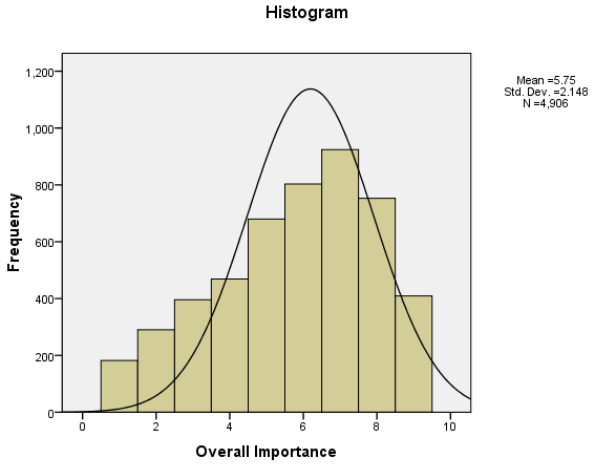
Q4. Please rate the utility of the health state **Deep Venous Thrombosis (DVT) in the Upper Leg – Severe**.
In the VAS scale below, you can either move the horizontal slider to your rating score or type your rating score in the box on the right. In this scale, a score of 100 represents "Full Health" and a score of 0 represents "Dead".

0 50 100

Chapter 3: Appendix F: Additional Data for Question Prioritization Regression Analysis

Figure 1: Histogram and Normal QQ Plot

Histogram and Q-Q plot showing left (negative) skew for dependent variable (overall importance).



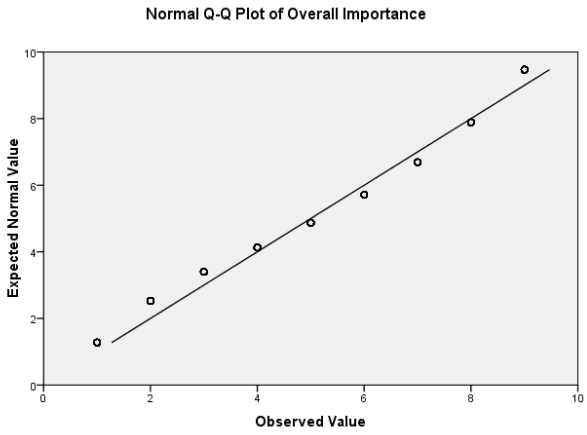
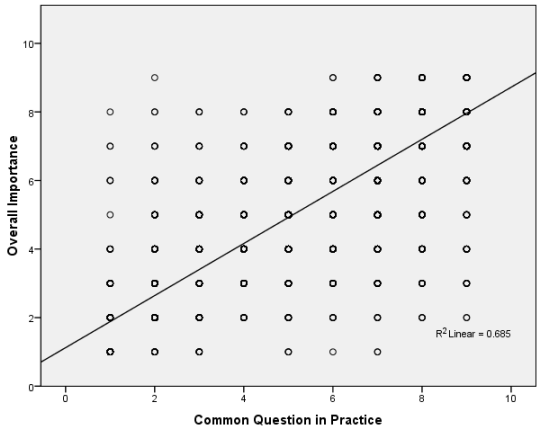
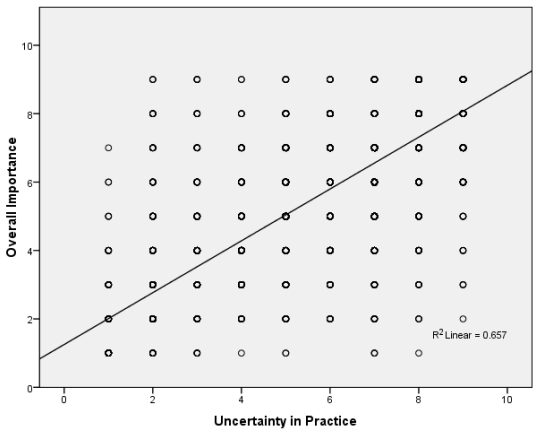
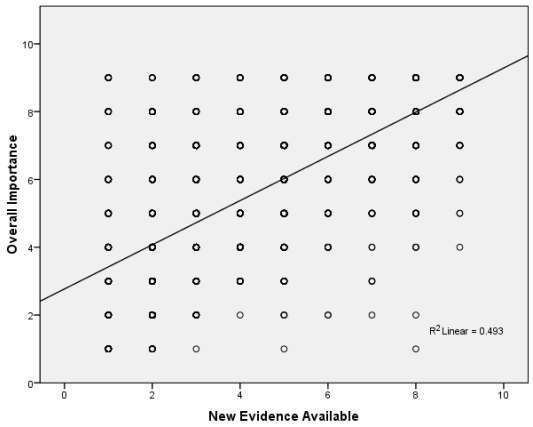


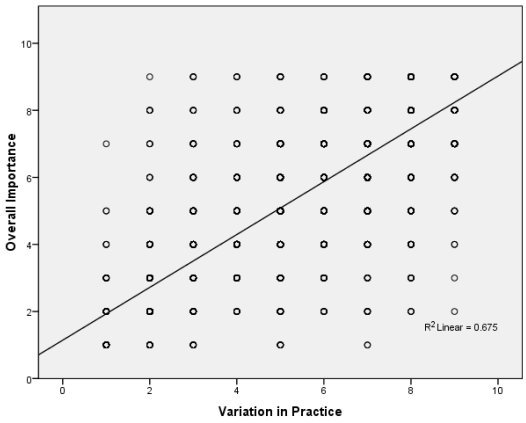
Figure 2: Scatterplots

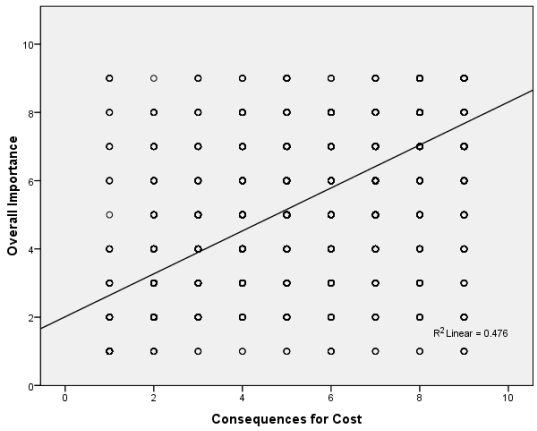
Scatterplots for each predictor variable vs. overall importance.






Overall Importance
New Evidence Available
R² Linear = 0.493


Overall Importance
Variation in Practice
R² Linear = 0.675


Overall Importance
Consequences for Cost
R² Linear = 0.476

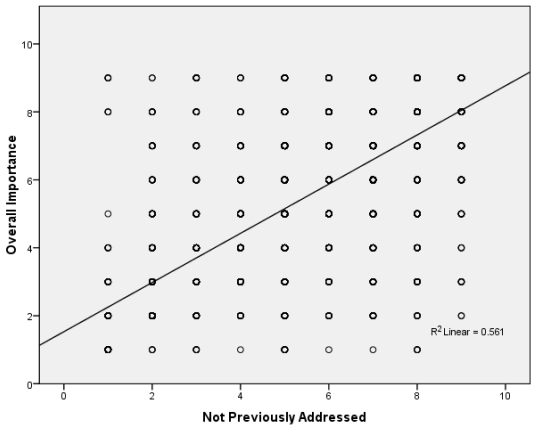

Overall Importance
Not Previously Addressed
R² Linear = 0.561

Table 1: Model Parameters and Estimates

Univariable Analysis				
Model	**Parameter**	**Estimate (β)**	***p*-value**	**95%CI**
1 - Random Intercept	Common Question in Practice	0.75	<0.01	0.73 to 0.76
2 - Random Intercept	Uncertainty in Practice	0.73	<0.01	0.71 to 0.75
3 - Random Intercept	New Evidence Available	0.68	<0.01	0.65 to 0.70
4 - Random Intercept	Variation in Practice	0.77	<0.01	0.75 to 0.78
5 - Random Intercept	Cost Consequences	0.64	<0.01	0.62 to 0.66
6 - Random Intercept	Not Previously Addressed	0.70	<0.01	0.68 to 0.73

Adjusted Analyses								
Model	**Parameter**	**Estimate of Fixed Effects (β)**	***p*-value**	**95%CI**	**Number of Parameters**	**-2 Log Likelihood**	**Aikeke's Information Criterion (AIC)**	**Variance of Residual**
Random Coefficients	Intercept	-0.19	0.04	-0.37 to -0.01	36	7658	7730	0.42
	Common Question in Practice	0.29	<0.01	0.24 to 0.34				
	Uncertainty in Practice	0.12	<0.01	0.09 to 0.15				
	New Evidence Available	0.14	<0.01	0.11 to 0.17				
	Variation in Practice	0.15	<0.01	0.11 to 0.19				
	Cost Consequences	0.12	<0.01	0.10 to 0.15				
	Not Previously Addressed	0.20	<0.01	0.16 to 0.24				

CHAPTER 4. ASSESSING THE PROCESS AND OUTCOME OF THE DEVELOPMENT OF PRACTICE GUIDELINES AND RECOMMENDATIONS: PANELVIEW INSTRUMENT DEVELOPMENT

Assessing the process and outcome of the development of practice guidelines and recommendations: PANELVIEW Instrument Development

Authors

Wojtek Wiercioch MSc [1]
Elie A. Akl MD, PhD [1,2]
Nancy Santesso MLIS, PhD [1]
Yuan Zhang PhD [1]
Rebecca L. Morgan MPH, PhD [1]
Juan José Yepes-Nuñez MD, PhD [3]
Sérgio Kowalski MD, PhD [4]
Tejan Baldeh, MPH [1]
Reem A. Mustafa MD, MPH, PhD [1,5]
Kaja-Triin Laisaar MD, MPH, PhD [6]
Ulla Raid MSc [7]
Itziar Etxeandia-Ikobaltzeta PharmD [1]
Alonso Carrasco-Labra DDS, MSc, PhD [1]
Matthew Ventresca MSc, MBA [1]
Ignacio Neumann MD, PhD [8]
Maicon Falavigna MD, PhD [9]
Romina Brignardello-Petersen DDM, PhD [1]
Gian Paolo Morgano MSc, PhD [1]
Jan Brożek MD, PhD [1, 10]
Meghan McConnell PhD [11]
Holger J. Schünemann MD, PhD [1,10]

for the *PANELVIEW Working Group:*

Yngve Falck-Ytter, MD [12,13]
Iván D. Florez, MD, MSc, PhD [1,14]
Amir Qaseem, MD, MHA, PhD [15]
Richard M. Rosenfeld MD, MPH, MBA [16]
Craig W. Robbins MD, MPH [17,18]
Judith Thornton, PhD [19]

Corresponding author: Holger J. Schünemann (schuneh@mcmaster.ca)

Author Affiliations:

1. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
2. Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
3. School of Medicine, Universidad de los Andes, Bogotá, Colombia
4. Department of Internal Medicine, Federal University of Paraná, Brazil.

5. Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, Kansas City, Kansas, USA
6. Guideline Development Group, Faculty of Medicine, Institute of Family Medicine and Public Health, University of Tartu, Estonia
7. Ministry of Social Affairs, Estonia
8. Department of Internal Medicine, Pontificia Universidad Católica de Chile, Chile
9. Federal University of Rio Grande do Sul, National Institute for Health Technology Assessment, Brazil
10. Department of Medicine, McMaster University, Hamilton, Ontario, Canada
11. Department of Innovation in Medical Education, University of Ottawa, Ottawa, Ontario, Canada
12. Division of Gastroenterology, Case Western Reserve University, Cleveland, Ohio, USA
13. Veterans Affairs Medical Center and University Hospitals of Cleveland, Cleveland, Ohio, USA
14. Department of Pediatrics, Universidad de Antioquia, Medellin, Colombia.
15. American College of Physicians, Philadelphia, Pennsylvania, USA
16. Department of Otolaryngology, SUNY Downstate Medical Center, Brooklyn, New York, USA
17. Center for Clinical Information Services, Kaiser Permanente Care Management Institute, Oakland, California, USA
18. Department of Health Systems Science, Kaiser Permanente School of Medicine, Pasadena, California, USA
19. Centre for Clinical Practice, National Institute for Health and Clinical Excellence, Manchester, United Kingdom

Funding statement: No external funding was utilized. Internal funds, McMaster GRADE centre, McMaster University.

Competing interests: Authors of this manuscript have been involved in the development of various guideline manuals which are referenced in this article. Three of the authors (WW, EAA, HJS) will retain copyright of the tool through McMaster University (Technology ID 20-031) and the tool may be licensed in the future. Maicon Falavigna declares receiving consulting fees from Novartis, Abbvie, Roche, Boehringer Ingelheim, PTC Therapeutics and Sanofi-Genzyme unrelated to the submitted work.

Abstract

Background: Guideline recommendations may be impacted by flaws in the process, inappropriate panel member selection or conduct, conflicts-of-interest, and other factors. Currently, no validated tool exists to evaluate guideline development from the perspective of those directly involved in the production. Our objective was to develop and validate a universal tool to assess guideline processes, methods and outcomes by the participating guideline panelists and group members.

Methods: A systematic literature search and surveys of guideline groups, identified through contacting international organizations and convenience sampling of working panels, informed item generation. Subsequent groups of guideline methodologists and panelists reviewed items for face validity and missing items. We used surveys, interviews, and expert review for item reduction and phrasing. For reliability assessment and feedback, we tested the PANELVIEW tool in eight international guideline groups.

Results: We surveyed 62 members from 13 guideline panels, contacted 19 organizations, and reviewed 20 source documents to generate items. Fifty-three additional key informants provided feedback about the items and response option phrasing. We reduced the PANELVIEW tool from 95 to 34 final items across domains that include administration, training, conflicts-of-interest, group dynamics, chairing, evidence synthesis, formulating recommendations, and publication. The tool takes approximately 10 minutes to complete and showed acceptable measurement properties.

Interpretation: The PANELVIEW tool enables guideline organizations to directly involve clinicians, patients and other participants in evaluating their guideline processes. The tool can inform quality improvement of existing or new guideline programs, focusing on insight into and transparency of the guideline development process, methods and outcomes.

Introduction

As a product of a group process that involves project planning, synthesizing evidence, and deliberation by guideline group members to reach consensus and formulate recommendations, health guidelines are highly influential in determining practice.(1, 2) Guideline development also requires careful coordination of multiple teams with specialized knowledge.(1, 3, 4) These teams typically include an oversight committee responsible for project planning, working groups responsible for preparation and technical aspects of evidence synthesis, and a guideline panel tasked with prioritizing questions and formulating recommendations.

Evidence suggests that guideline group processes may be prone to influence by individuals with strong opinions, imbalanced group member characteristics, unqualified members, or not using the best available evidence.(2, 5-8) Currently available instruments assessing the trustworthiness of practice guidelines rely on what the guideline authors report, typically in peer reviewed publications, or from reports of organizations or manuals.(9, 10) However, what authors report may be generic or incomplete, lack transparency, or be inconsistent with the assessment of all group members, and what is reported may not always reflect what happened.(11) For example, the AGREE(9) and RIGHT(10) tools appropriately call for conduct of systematic reviews as part of guideline development and appropriate disclosure of potential influence of conflicts. While systematic reviews should inform guidelines, their conduct does neither guarantee their quality nor that they are used appropriately by the panel

for making recommendations. Likewise, having a conflict-of-interest declaration and management policy for guidelines does not necessarily guarantee that conflicts are well-managed when guideline groups make recommendations.

Existing tools do not evaluate essential steps and processes as they take place, such as giving appropriate consideration to the evidence and ensuring that all panel members have an equal voice.(9, 10, 12) An internal evaluation by participating guideline group members would provide this valuable insight. Additionally, ensuring that panel members view the process as appropriate and one that results in a credible guideline will help ensure they view value in their contribution. By obtaining an assessment from the participants, guideline developers could identify areas of their processes viewed as needing improvement, identify dissenting views amongst participants, use this information to modify their methods and approaches and ensure the credibility of their guidelines and trustworthiness of recommendations.

The objective of this research was to develop and validate the PANELVIEW instrument for assessing guideline panel members' perception of the appropriateness of, and satisfaction with, the process, methods and outcome of the development of a health guideline. Organizations responsible for guideline development can use the PANELVIEW tool by asking the participating group members who serve as clinicians, patient representatives, content experts, methodologists, and other stakeholders.

Methods

For the development of the PANELVIEW tool we followed methods for scale development, including generation of items based on existing literature, item reduction through key informant and expert feedback and consensus, and field testing with guideline panels (see Figure 1).(13)

Item Generation

Item generation began with two investigators (WW, HJS) discussing key domains for capturing the evaluation of guideline-related processes based on domains in the GIN-McMaster Guideline Development Checklist.(1) We hypothesized that all parts of the process might be relevant for assessing appropriateness and satisfaction of panel members. We then conducted a systematic literature search to identify steps and themes in guideline development that relate to the appropriateness of the process. This step was followed by key informant surveys and interviews with panelists participating in guidelines and guideline methodologists. We recruited key informants through convenience sampling, informed by contacting guideline organizations and through the study team to identify working panels. For each step involving key informants, we used a new sample of participants as a method of confirming data and views obtained in the preceding step.

Literature Review

The systematic search of the literature in Medline and Embase (from inception to November 2018) aimed to identify studies that discussed or evaluated steps of guideline

development.(1) We used controlled vocabulary and keywords to capture evaluation of the guideline development process and panel member perceptions (see Appendix Figure 1 and Box 1 for additional details). To supplement the literature search, we contacted a convenience sample of 19 key informants, identified in a previous project,(1) representing major guideline-developing organizations globally. We asked if the organizations currently conduct internal evaluation of their guideline development processes or use specific tools. (see Figure 1 Step 1).

Panel Surveys

We surveyed members from 13 guideline panels to obtain primary data (see Figure 1 Step 1). Sixty-two panelists completed hardcopy surveys when their meetings adjourned, consisting of six open-ended questions inquiring about the factors that impacted their satisfaction and perception of the appropriateness of the process (see Appendix Table 1 and Figure 2). The survey responses were included as a source document for data abstraction.

Data Abstraction from Item Generation Sources

We developed and pilot tested a structured data abstraction form. Study team members (YZ, RM, KTL, UR, MV, JJYN, RM, NS, SK, TB) reviewed full texts of source documents and abstracted independently and in duplicate items that related to the appropriateness of guideline development methods or processes, panel members' views about methods or processes, or panel members' satisfaction. Supporting quotations from the source document were included for each item, along with proposed themes

that an item could be grouped into (e.g., conflict of interest management, training, group interaction, etc.).

A subgroup of the study team members (WW, YZ, RM, SK, JJYN) independently identified and merged duplicate items (i.e. measuring or asking about the same aspect of the guideline development process). The decisions were assessed by a second reviewer and discussed in a team meeting during which we finalized the de-duplication, initial item phrasing, and allocation to specific themes of the guideline development process.

Item Reduction

Feedback from Key Informants

We surveyed and conducted interviews with a convenience sample of 22 key informants, including guideline developers, methodologists, and panel members to obtain feedback about the initial list of items (see Figure 1 Step 2 and Appendix Table 2). Respondents were asked to rate on a 7-point Likert-type scale (1-not important; 7-very important) how important they considered each of the items to be for evaluating the guideline development process, to suggest modifications, and to identify any missing items. We also sought to obtain in-depth feedback about the initial list and asked participants to comment on the level of detail, clarity, and redundancy in the items. We then sent the list of items to participants for review in advance and conducted interviews in person at guideline panel meetings in presence of a note-taker. We pilot tested both the survey and interview guide.

Study Team Review and Consensus Meeting

Concurrently with the key informant surveys and interviews, we provided study team members (NS, IEI, YZ, KTL, SK, RBP, MV, MF, GPM) with a structured feedback form to also review the initial list of items and provide suggestions for modifications, for the theme categorization, or to suggest potentially missing items. For each item, we summarized the study team's suggestions and the key informants' feedback and rating of importance. Study team members reviewed the summary and individually suggested to either keep, modify, merge, or delete items in preparation for a consensus meeting. We finalized decisions about each item based on discussion and group consensus and we refined item wording based on our experience with developing other measurement instruments.(14, 15)

Item and Response Option Phrasing

We conducted surveys with 26 additional key informants to determine the phrasing of items and response options (see Figure 1 Step 3 and Appendix Table 2). We asked respondents to choose their preference for one of three 7-point Likert-type response and item phrasing options, presenting eight example items from the tool. The first option asked about appropriateness, the second about satisfaction, and the third, representing the original Likert scale, about agreement with the topic presented by the item.

Testing with Panels

After item reduction, we pilot tested the PANELVIEW tool with 1 guideline panel and then used it with an additional 8 guideline panels (see Figure 1 Step 4 and Appendix Table 3). Panel members completed the PANELVIEW survey individually, expressing their agreement on the 7-point Likert scale with each item (1-strongly disagree; 7-strongly agree) (e.g. "There was appropriate management of potential bias in panel members' interpretation of evidence and alignment with prior beliefs"). Panel members also provided feedback on the clarity of the instructions, clarity of items and the survey length.

Analysis

We used generalizability theory (G-theory) to assess the reliability of scores obtained across the different panel groups.(13) We calculated the item mean scores, standard deviations, and ranges across individual panelists. For individual panelists, the overall scores were calculated as the mean of their item ratings. We conducted the preliminary reliability analyses at the individual panelist level, and at the panel level, whereby item means were obtained by collapsing across individual panelists. We estimated multiple sources of variance (*G)*, including the respondent, panel, item, and domain using a nested G-theory study.(13) The guideline development process of different panel groups served as the object of measurement, individual respondents were nested within panels, and individual items were nested within the PANELVIEW survey domains (see Appendix Table 5 and Figure 5).

IRB Approval

The Hamilton Integrated Research Ethics Board approved this study prior to data collection (Project #14-867).

Results

Item Generation

Our systematic literature search, contact with key informants and surveys of 13 guideline panels yielded 17 published articles,(6, 16-31) 3 additional source documents (see Appendix Figure 3) and 62 survey responses. We abstracted a list of 694 items, which after evaluation and de-duplication resulted in 95 items grouped across 17 themes covering guideline development (see Appendix Table 4).

Item Reduction and Phrasing of the Response Options

Informed by the rating of importance of the 95 items and feedback from key informants, we removed 23 items that scored low on importance as part of our consensus process. Thirty-eight items were merged with other items considered redundant. We phrased each item to ensure it assessed only one component of the guideline development process. The final, reduced list included 34 items. The feedback from the 26 key informants about the phrasing of response options indicated preference (73%) on the Likert scale.

Testing with Panels

After pilot testing with the 1 panel consisting of 12 panel members, we made minor revisions to clarify item wording and the order of items in the tool. We then obtained responses from 94 panelists through field testing with the 8 guideline panels (see Appendix Table 3).

Generalizability and reliability

The analysis of variance from the nested *G*-study showed an overall test reliability coefficient of 0.35 (see Appendix Table 5). This result is likely an effect of enrolling homogenous panels with regards to processes and methods. The tool's domains and individual items within the domains accounted for 4% of the variance, respectively, also suggesting that the processes for the guideline efforts we evaluated were similar across the domains and items. The guideline panels accounted for 28% of the variance and participants within panels accounted for 55% of the variance in scores. This indicates that variation was captured in panelists' assessments between the guideline panels and in panelists' ratings within the panels. Despite the similarity across groups, the tool was able to identify varying views of guideline panel members indicating higher and lower satisfaction or perception of appropriateness.

Response variation, item-item correlation and internal consistency

Within the panels, item means ranged from 4.0 to 7.0 and the item-item correlations ranged from -0.76 to 0.96, while item-total correlations ranged from -0.17 to 0.89. Across the 8 panels the item means ranged from 5.5 to 6.8 (see Table 1). There

was high internal consistency in rating of satisfaction and appropriateness of the process within the 8 panels, with Cronbach's α ranging from 0.85 to 0.98 (see Table 2). For individual panelists, we found item responses ranged from 1 to 7 on the Likert scale and the item-item correlations ranged from 0.003 to 0.719. This suggests, on an individual respondent level, that the tool distinguishes between responses and that there is no end-of-scale aversion. Item-total correlations by individual raters ranged from 0.40 to 0.80 which suggests that the items are measuring different aspects of the guideline process.

Feedback from guideline panel group members

Respondents agreed on the Likert scale that they did not have difficulty completing the questionnaire, with a mean rating of 6.4 (SD 0.6). Respondents, on average, felt that the questionnaire was neither too long nor too short (mean rating of 3.5, SD 1.7, with a score below 4 suggesting that the questionnaire is not too long). We observed an average time to complete the survey of 12 minutes (SD 7 minutes) and a median time of 10 minutes for 68 respondents completing it online. This estimate excluded 12 respondents with a recorded completion time of 30 minutes or longer, who presumably took a break while completing the questionnaire (see Appendix Figure 4). Eight respondents suggested to provide an option to respond to items as 'not applicable', which we added to relevant items (e.g. to allow panel chairs to skip items that request evaluation of their chairing of the panel). The final PANELVIEW tool is available at https://heigrade.mcmaster.ca/guideline-development/panelview.

Interpretation

We developed a tool that allows guideline developers to assess their processes, methods and outcomes by directly involving clinicians, patients and any other guideline group member in the evaluation. We followed best practice for instrument development, including reviewing the literature, contacting key informants at guideline organizations, and surveying panelists about key factors impacting guideline development. We successfully tested the tool with panels from guideline organizations globally.

Use of PANELVIEW

The tool enables evaluation of guideline development by participating group members in its entirety or in phases. How the guideline process is organized may differ between organizations, for example between those that convene one final panel meeting and those that maintain a standing panel with repeated meetings. This will determine whether developers administer the PANELVIEW evaluation once at the conclusion of a guideline project, or throughout the process as the steps take place. The objective of the tool is to identify strengths and weaknesses of a guideline development group's process and methods in a structured manner, and highlight specific areas for improvement as identified by the participants by assessing ratings within individual domains.

Implications for Practice

The tool is not intended to replace existing tools that offer guidance on the appropriate steps for guideline development or assess the credibility of published guideline reports. It offers a novel approach for identification of issues in the guideline development process and methods by those who participate in it or directly observe it, such as technical experts and methodologists. The tool can serve to inform evaluation or quality improvement of new or existing guideline programs, respectively.

Strengths and Limitations

The rigor of development with the end user in mind is the main strength of our work. First, we applied item generation methods drawing on multiple sources: literature, contacting organizations, panel surveys, and a team with extensive experience in the guideline field. Second, we involved other key informants from multiple organizations and participation on panels for input on items and face validity, allowing data saturation. Third, we field tested the tool with groups focusing on a variety of guideline topics.

A potential limitation of our research is that we did not conduct systematic searches of the non-medical literature in the areas of business, education, and policy-making for relevant items. At each step involving key informants we used convenience sampling, which may introduce a sampling bias. To address this, we drew on a broad representation of working guideline panelists, with varying levels of experience, as well as guideline development experts from organizations representing a wide range of processes and methods.

In field testing the PANELVIEW tool, the 8 guideline groups were recruited through key informants and for some aspects of development the groups used similar methods (e.g. using the GRADE approach for assessing quality of evidence and strength of recommendations) and involved experienced group chairs. The high scores on many items and the lower overall reliability coefficient of 0.35 to discriminate between groups indicated that the groups were likely all highly performing. Despite this, we observed variability in scores within the groups, which would allow guideline developers to identify whether individual panelists viewed the process and specific aspects of the process as more or less appropriate.

Next Steps and Future Research

In the next steps, we seek to administer the PANELVIEW tool with additional, diverse panels from various guideline organizations. Guideline organizations can access the tool at https://heigrade.mcmaster.ca/guideline-development/panelview to participate. We will seek further feedback on use of the tool, for example about the potential for public reporting of PANELVIEW assessments to increase transparency. The high Cronbach alpha coefficients may indicate the presence of redundant items. Sampling of more panels will allow us to assess if any necessary refinement of tool items is necessary and conduct factor analysis for further evaluation of tool domains. Additional opportunities include comparative studies, for example using PANELVIEW assessments of panelists with that of other group members (e.g. non-voting observers), as well as

evaluating global ratings and judgements of panel success and guideline credibility against ratings of the tool.

Conclusion

We believe that PANELVIEW addresses a critical area in guideline development as no existing and validated tool exists that allows capturing of panelists' perspectives when they participate in guidelines. Existing instruments for assessing guideline credibility rely on the guideline authors' report, which may describe the process as planned but not as implemented or as viewed by all group members, and may not reflect all relevant nuances of the process that impact the trustworthiness of recommendations. The PANELVIEW tool focuses on these important nuances and transparency of the guideline development process, allowing organizations responsible for guideline development to inform their quality improvement efforts. Given the importance of guidelines and their impact on recipients and providers of care, optimizing the quality of their development is a logical step.

Affiliations: Department of Health Research Methods, Evidence, and Impact (Wiercioch, Akl, Santesso, Zhang, Morgan, Mustafa, Etxeandia-Ikobaltzeta, Carrasco Labra, Ventresca, Brignardello-Petersen, Morgano, Brożek, Florez, Schünemann), and Department of Medicine (Brożek, Schünemann), McMaster University, Canada; Department of Internal Medicine (Akl), American University of Beirut, Lebanon; School of Medicine (Yepes-Nuñez), Universidad de los Andes, Bogotá, Colombia; Department of Internal Medicine (Kowalski), Federal University of Paraná, Brazil; Division of Nephrology and Hypertension, Department of Medicine (Mustafa), University of Kansas Medical Center, United States; Guideline Development Group, Faculty of Medicine, Institute of Family Medicine and Public Health (Laisaar), University of Tartu, Estonia; Ministry of Social Affairs (Raid), Estonia; Department of Internal Medicine (Neumann), Pontificia Universidad Católica de Chile, Chile; Federal University of Rio Grande do Sul, National Institute for Health Technology Assessment (Falavigna), Brazil; Department of Innovation in Medical Education (McConnell), University of Ottawa, Canada; Division of Gastroenterology (Falck-Ytter), Case Western Reserve University, United States; Veterans Affairs Medical Center and University Hospitals of Cleveland (Falck-Ytter), United States; Department of Pediatrics (Florez), Universidad de Antioquia, Medellin, Colombia; American College of Physicians (Qaseem), United States; Department of Otolaryngology (Rosenfeld), SUNY Downstate Medical Center, United States; Center for Clinical Information Services (Robbins), Kaiser Permanente Care Management Institute, United States; Family Medicine (Robbins), Colorado Permanente Medical Group, United States; Centre for Clinical Practice, National Institute for Health and Clinical Excellence (Thornton), United Kingdom.

Contributors: Holger Schünemann was the principal investigator, had full access to the study data and takes responsibility for the integrity of the data and the accuracy of the data analysis. Holger Schünemann, Elie Akl, and Wojtek Wiercioch were responsible for conceptualization of the research study. All authors contributed to the acquisition and interpretation of data. Meghan McConnell and Wojtek Wiercioch were responsible for carrying out the data analysis. Wojtek Wiercioch and Holger Schünemann drafted the manuscript; all of the authors revised the manuscript critically for important intellectual content and approved the final version submitted for publication.

Funding: This study was supported by internal funds from the McMaster GRADE centre, McMaster University.

Kaja-Triin Laisaar receives institutional research support through grant IUT34-17 “Health research in a continuum of evidence-based health practice in Estonia” from the Ministry of Education and Research through the Estonian Research Council.

Acknowledgements: The authors would like to acknowledge Dr. John J. Riva for helping to pilot test the item reduction survey and Dr. Gordon Guyatt for helping to field test the PANELVIEW tool. We also wish to acknowledge all the participating guideline panel members who provided feedback for item generation and reduction and helped with field testing of the PANELVIEW tool.

References

1. Schünemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. Canadian Medical Association Journal. 2014;186(3):E123-E42.
2. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 5. Group processes. Health Research Policy and Systems. 2006;4(1):17.
3. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011.
4. Qaseem A, Forland F, Macbeth F, Ollenschlager G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156(7):525-31.
5. Schunemann HJ, Woodhead M, Anzueto A, Buist AS, Macnee W, Rabe KF, et al. A guide to guidelines for professional societies and other developers of recommendations: introduction to integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proceedings of the American Thoracic Society. 2012;9(5):215-8.
6. Neumann I, Karl R, Rajpal A, Akl EA, Guyatt GH. Experiences with a novel policy for managing conflicts of interest of guideline developers: a descriptive qualitative study. Chest. 2013;144(2):398-404.
7. Boyd E, Bero L. Improving the use of research evidence in guideline development: 4. Managing conflicts of interests. Health Research Policy and Systems. 2006;4(1):16.
8. Fretheim A, Schunemann H, Oxman A. Improving the use of research evidence in guideline development: 3. Group composition and consultation process. Health Research Policy and Systems. 2006;4(1):15.
9. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. Canadian Medical Association Journal 2010;182(18):E839-42.
10. Chen Y, Yang K, Marušić A, et al. A reporting tool for practice guidelines in health care: The right statement. Annals of Internal Medicine. 2017;166(2):128-32.
11. Alexander PE, Li SA, Gionfriddo MR, Stoltzfus RJ, Neumann I, Brito JP, et al. Senior GRADE methodologists encounter challenges as part of WHO guideline development panels: an inductive content analysis. Journal of clinical epidemiology. 2016;70:123-8.
12. Shiffman RN, Shekelle P, Overhage JM, Slutsky J, Grimshaw J, Deshpande AM. Standardized reporting of clinical practice guidelines: a proposal from the Conference on Guideline Standardization. Ann Intern Med. 2003;139(6):493-8.

13. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. USA: Oxford University Press; 2015.
14. Puhan MA, Behnke M, Laschke M, Lichtenschopf A, Brandli O, Guyatt GH, et al. Self-administration and standardisation of the chronic respiratory questionnaire: a randomised trial in three German-speaking countries. Respir Med. 2004;98(4):342-50.
15. Schunemann HJ, Goldstein R, Mador MJ, McKim D, Stahl E, Puhan M, et al. A randomised trial to evaluate the self-administered standardised chronic respiratory questionnaire. Eur Respir J. 2005;25(1):31-40.
16. Atkins L, Smith JA, Kelly MP, Michie S. The process of developing evidence-based guidance in medicine and public health: a qualitative study of views from the inside. Implement Sci. 2013;8:101.
17. Calderon C, Rotaeche R, Etxebarria A, Marzo M, Rico R, Barandiaran M. Gaining insight into the Clinical Practice Guideline development processes: qualitative study in a workshop to implement the GRADE proposal in Spain. BMC Health Services Research. 2006;6:138.
18. Chang LW, Kennedy CE, Kennedy GE, Lindegren ML, Marston BJ, Kaplan JE, et al. Developing WHO guidelines with pragmatic, structured, evidence-based processes: A case study. Global Public Health. 2010;5(4):395-412.
19. Gardner B, Davidson R, McAteer J, Michie S. A method for studying decision-making by guideline development groups. Implement Sci. 2009;4:48.
20. Harding E, Brown D, Hayward M, Pettinari CJ. Service user perceptions of involvement in developing NICE mental health guidelines: A grounded theory study. Journal of Mental Health. 2010;19(3):249-57.
21. Hopthrow T, Feder G, Michie S. The role of group decision making processes in the creation of clinical guidelines. Int Rev Psychiatry. 2011;23(4):358-64.
22. Hutchings A, Raine R. A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. Journal of Health Services & Research Policy. 2006;11(3):172-9.
23. Hutchings A, Raine R, Sanderson C, Black N. An experimental study of determinants of the extent of disagreement within clinical guideline development groups. Quality & Safety in Health Care. 2005;14(4):240-5.
24. Khodyakov D, Hempel S, Rubenstein L, Shekelle P, Foy R, Salem-Schatz S, et al. Conducting online expert panels: a feasibility and experimental replicability study. BMC Med Res Methodol. 2011;11:174.
25. Moreira T, May C, Mason J, Eccles M. A new method of analysis enabled a better understanding of clinical practice guideline development processes. Journal of clinical epidemiology. 2006;59(11):1199-206.
26. O'Brien KK, Wilkins A, Zack E, Solomon P. Developing clinical practice guidelines in HIV rehabilitation: process recommendations and guiding principles. AIDS Educ Prev. 2011;23(5):457-68.

27. Pagliari C, Grimshaw J. Impact of group structure and process on multidisciplinary evidence-based guideline development: an observational study. J Eval Clin Pract. 2002;8(2):145-53.
28. Pagliari C, Grimshaw J, Eccles M. The potential influence of small group processes on guideline development. Journal of Evaluation in Clinical Practice. 2001;7(2):165-73.
29. Raine R, Sanderson C, Black N. Developing clinical guidelines: a challenge to current methods. Bmj. 2005;331(7517):631-3.
30. Raine R, Sanderson C, Hutchings A, Carter S, Larkin K, Black N. An experimental study of determinants of group judgments in clinical guideline development. Lancet. 2004;364(9432):429-37.
31. Williams W, Lowery NE, Lyalin D, Lambrecht N, Riddick S, Sutliff C, et al. Development and utilization of best practice operational guidelines for immunization information systems. Journal of Public Health Management & Practice. 2011;17(5):449-56.

Chapter 4: Tables and Figures

Figure 1: Overview of steps and participants in the PANELVIEW tool development

Tool Development Step	Methods	Sources & Participants	PANELVIEW Items
1. Item Generation	Systematic Review	17 articles	694
	Contacting Guideline Developers	3 source documents from 19 developers (AAO-HNS, ACP, CAR, CCO, CDC, Chile MoH, Colombia MoH, CTFPHC, DCGP, FMSD, KP, NHMRC, NICE, NKCHS, OPHA, RKIG, SIGN, South Africa MoH, USPSTF)	
	Key Informant Survey	62 panelists: 13 panels (WHO and KSA MoH)	
2. Item Reduction	Key Informant Survey	9 panelists and guideline methodologists: 3 panels (WHO, WAO, CCO)	95
	Key Informant Interview	13 panelists and guideline methodologists: 2 panels (WAO, Estonia MoH)	
	Expert Review and Consensus	Study team experts and guideline methodologists	
3. Response Phrasing	Key Informant Survey	26 panelists and guideline methodologists: 3 workshops, 1 panel (AGA)	34
4. Field Testing	Use of Tool with Working Panels	12 panelists, 1 pilot panel (NHF) 94 panelists, 8 panels (AABB, ADA, RA Adaptation, RARE-Bestpractices, WHO)	

Abbreviations:
AAO-HNS - American Academy of Otolaryngology–Head and Neck Surgery; ACP - American College of Physicians; ADA – American Dental Association; AGA - American Gastroenterological Association; CAR - Canadian Association of Radiologists; CCO – Cancer Care Ontario; CDC - Centers for Disease Control and Prevention; CTFPHC - Canadian Task Force on Preventive Health Care; DCGP - Dutch College of General Practitioners; FMSD - Finnish Medical Society Duodecim; KP - Kaiser Permanente; KSA – Kingdom of Saudi Arabia; MoH – Ministry of Health; NHMRC - National Health and Medical Research Council Australia; NICE - National Institute for Health and Care Excellence; NKCHS - Norwegian Knowledge Centre for the Health Services; OPHA - Ontario Public Health Agency; RA – Rheumatoid Arthritis; RKIG - Robert Koch Institute Germany; SIGN - Scottish Intercollegiate Guidelines Network; USPSTF - U.S. Preventive Services Task Force; WAO – World Allergy Organization; WHO – World Health Organization

Figure 1 Caption: The PANELVIEW tool development consisted of 4 main steps. At each step information about the items to evaluate the appropriateness of guideline development was evaluated with participating panel members and guideline methodologists drawn from organizations with representation of diverse geographic and clinical topic areas. The key informants for each step were drawn from new samples to ensure broad representation of views and perspectives.

Table 1: PANELVIEW tool mean scores across panels

Domain	Item	Mean Score	Standard Deviation	Min-Max
1. Administration	1 - The logistical support provided for organization of the guideline project and panel meeting was appropriate (e.g. scheduling of meeting, sharing of materials, venue/location).	6.29	0.86	3-7
	2 - There was adequate preparatory work and meetings/teleconferences prior to the final panel meeting.	5.79	1.34	2-7
	3 - Adequate time was given for guideline group members to complete tasks (e.g. surveys, providing feedback) throughout the development of the guideline and to review the evidence summary and other material prior to the panel meeting.	5.87	1.12	2-7
	4 - Adequate time was allotted for the final panel meeting for all guideline questions to be discussed and recommendations to be formulated.	5.54	1.49	2-7
	5 - The panel meeting had a clearly defined agenda and objectives.	6.48	0.87	3-7
2. Training	6 - Information was provided about the specific methodology and frameworks to ensure understanding of the overall process and steps that would be used to develop the guideline	6.35	0.81	3-7
3. Panel Chair	7 - The panel chair(s) was able to provide clinical and methodological guidance during the meeting, providing direction and support for decision-making	6.56	0.78	3-7
	8 - The panel chair(s) was able to manage the group process, establishing an atmosphere of support that ensured involvement of all panel members in the discussion and free expression of opinions	6.60	0.63	5-7
4. Conflict of Interest	9 - There was appropriate management of potential interests (financial, academic) of guideline group members, of the organization, and in the evidence synthesis being free from bias	6.23	1.11	1-7
	10 - There was appropriate management of potential bias in panel members' interpretation of evidence and alignment with prior beliefs	5.97	1.20	1-7
5. Scoping the Guideline	11 - The panel was given sufficient opportunity to be involved in the prioritization of questions and scoping of the guideline	6.26	0.84	4 -7

	12 - The final scope of the guideline was clearly communicated to the guideline group and agreement was sought	6.32	0.83	4-7
6. Methodology and Process	13 - The evidence synthesis was rigorous	6.14	0.97	3-7
	14 - A transparent and usable summary of the evidence was made available for the discussion	6.23	1.14	2-7
7. Considering the Evidence and Contributing through Expertise	15 - Appropriate consideration was given to the evidence, including all relevant types, and balanced with panel members' input and opportunity to use their experience to interpret the evidence	6.45	0.71	3-7
	16 - The method or process used for decision making with the available evidence was appropriate	6.40	0.74	3-7
	17 - There was appropriate involvement and consultation with key stakeholders during the guideline development	5.84	0.97	4-7
	18 - Appropriate consideration was given to patients' views, perspective, values and preferences	5.71	1.20	2-7
8. Formulating the Recommendations	19 - An appropriate method was used for formulating the recommendations with transparency of judgements made	6.45	0.77	3-7
	20 - Appropriate consideration was given to relevant external factors (e.g. policy implications, setting-specific healthcare factors, acceptability of recommendations) in formulating the guideline recommendations	6.12	0.88	2-7
	21 - The consensus method used by the panel was appropriate, allowing ability to reach consensus	6.36	0.72	4-7
	22 - The wording of the guideline recommendations formulated was clear and actionable	6.26	0.79	4-7
	23 - There was transparency in going from the panel's recommendation to the final recommendations that appear in the guideline report and notice was given about any changes made	6.24	0.99	4-7
9. Group Composition	24 - There was diversity in membership and adequate representation of backgrounds, specialties and balance of expertise in the panel composition	6.35	0.88	3-7
	25 - The panel size was appropriate	6.41	0.81	3-7

10. Group Roles	26 - The required commitment was at an appropriate level for the guideline group members	6.47	0.67	4-7
	27 - The contributions of the guideline group members were valued and appropriate credit was given	6.52	0.68	4-7
11. Group Interaction	28 - There was mutual respect between guideline group members with friendly and professional conduct	6.71	0.54	5-7
12. Implementation and Dissemination Planning	29 - Appropriate consideration was given to the discussion of research gaps and needs for future research	6.28	0.85	3-7
	30 - Appropriate consideration was given for the planning of dissemination and implementation of the guideline	6.08	1.04	3-7
13. Writing of the Guideline	31 - The writing of the guideline was well planned, with agreement on the format(s) and opportunity for panel members to provide input and review the guideline draft	5.90	1.19	2-7
14. Incentive	32 - I felt that my involvement in the guideline will have an impact on the health of people.	6.30	0.81	4-7
15. Overall Satisfaction	33 - Overall, I was satisfied with the guideline development process	6.48	0.69	4-7
	34 - I would participate in this guideline development process again	6.78	0.44	5-7

Table 2: PANELVIEW tool mean scores and internal consistency across guideline groups

Guideline Panel	Mean Score (SD)	Cronbach's α
1	6.46 (0.32)	0.92
2	6.04 (0.43)	0.98
3	6.53 (0.28)	0.88
4	6.05 (0.59)	0.96
5	6.27 (0.53)	0.95
6	6.07 (0.34)	0.96
7	6.37 (0.24)	0.95
8	6.01 (0.50)	0.85

Chapter 4: Appendix 1 - Tables

Appendix Table 1: Guideline groups and panelists involved in item generation surveys

Guideline Organization	Number of Respondents
Kingdom of Saudi Arabia Ministry of Health Guidelines 2014	38
World Health Organization 2013	10
World Health Organization 2014	11
ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis 2014	3

Appendix Table 2: Key informants involved in item reduction and item and response phrasing

Sampling	Number of Respondents	Previously Participated in a Guideline (%)	Field of Work*				
			Clinical (%)	Research (%)	Administrative (%)	Policymaking (%)	Teaching (%)
Item Reduction							
WHO, WAO, CCO, Estonia MoH panelists and methodologists	22	91	68	91	23	32	59
Item and Response Phrasing							
Guideline workshops and AGA panelists and methodologists	26	72	72	56	8	2	4

* participants were able to select more than one category

Abbreviations:
AGA - American Gastroenterological Association; CCO – Cancer Care Ontario; MoH – Ministry of Health; WAO – World Allergy Organization; WHO – World Health Organization

Appendix Table 3: Guideline panels involved in field testing the PANELVIEW tool

Guideline Organization	Guideline Topic	Date	Number of Respondents	Previously Participated in a Guideline (%)	Field of Work*				
					Clinical (%)	Research (%)	Administrative (%)	Policymaking (%)	Teaching (%)
National Hemophilia Foundation (*pilot guideline group*)	Care models for haemophilia management	July 7, 2015	12	33	67	67	42	17	50
AABB (*formerly American Association of Blood Banks)*	Red blood cell transfusion	January 7, 2016	14	93	86	64	29	21	36
American Dental Association	Sealants	January 22, 2016	8	38	88	75	63	25	63
Rheumatoid Arthritis Guideline Adaptation for the Eastern Mediterranean Region	Treatment of rheumatoid arthritis	May 27, 2016	17	47	94	65	47	0	59
RARE-Bestpractices	Sickle cell disease	July 11, 2016	8	75	75	88	38	0	63
McMaster RARE-Bestpractices	Catastrophic antiphospholipid syndrome	April 26, 2017	13	77	92	69	23	0	38
World Health Organization	Policy guidance on the use of delamanid in children	May 4, 2017	13	54	85	92	46	15	77
Rheumatoid Arthritis Guideline	Treatment of rheumatoid arthritis	July 7, 2017	12	58	100	67	50	0	83

Adaptation for the Eastern Mediterranean Region – Panel 2									
World Health Organization	Health workers guideline	December 15, 2017	9	11	33	44	56	56	67

* participants were able to select more than one category

Appendix Table 4: Initial list of items and domains prior to item reduction

Administration
1. Logistical support provided for organization for the panel meeting(s) (e.g. scheduling of meeting, setting agenda, booking travel, processing of expenses)
2. Planning, preparatory meetings, conference calls prior to final panel meeting(s)
3. Location and venue for panel meeting(s)
4. Adequate time given for guideline group members to complete tasks (e.g. completing surveys, providing input, etc.) throughout development of the guideline
5. Adequate duration of panel meeting(s) and time allotted for all guideline questions to be discussed and recommendations to be formulated
6. Materials being sent in advance with adequate time to review the evidence summary and other material prior to panel meeting
7. Panel meeting(s) have clearly defined objectives and agenda
8. The number of meetings held throughout the development of the guideline
Training
9. Training received about the specific methodology and frameworks to be used to develop the guideline in preparation for panel meeting(s)
10. The purpose and objectives of the entire guideline development project are clearly communicated to the guideline development group members
11. Information is provided to ensure understanding of the overall process and steps that will be used to develop the guideline
Panel Chair
12. Panel Chair's subject matter knowledge and expertise
13. Clear communication by panel Chair; easy to understand
14. Time management at the panel meeting(s) by the Chair; following agenda, staying on task and ensuring completion
15. Chair's ability to facilitate discussion, keeping discussion on topic, providing direction and support for decision-making, and maintaining fidelity of the process
16. Chair's ability to establish atmosphere of support that ensures involvement of all panel members in discussion and free expression of opinions
17. Chair's ability to manage group process and dynamics, and awareness of social, power, and knowledge influences in the group
18. Chair's ability to provide methodological guidance during panel meeting and adhere to the outlined methods and process
Conflict of Interest
19. Panel members completing Declaration of Interests (e.g. COI)
20. Management of potential conflicts of interest (financial, academic) and influence of networks that group members might mobilize during discussion
21. Management of bias in panel members' interpretation of evidence and alignment with prior beliefs
22. Independence of panel's decisions from the sponsoring guideline development organization's potential interests and influence
23. Evidence synthesis (e.g. systematic review) completed independently
Methodology & Process
24. Rigour of the evidence synthesis

25. Use of evidence in the formulation of recommendations for the guideline
26. Having specific procedures and methodology guiding the development of the guideline (e.g. as outlined in a handbook)
27. Adherence to the agreed on guideline development process and methods
28. Guideline development process and methods are transparent and communicated clearly to guideline group members
29. Involvement of panel members in evidence synthesis and contributing information
30. Involvement of and consultation with key stakeholders
Scoping the Guideline
31. Involvement of **all** guideline development group members in prioritization of questions and scoping of the guideline
32. The method used to decide on the scope of the guideline (e.g. literature search, rating exercise, stakeholder consultation)
33. Final scope of the guideline clearly communicated to the guideline development group and agreement sought
Considering the Evidence and Contributing through Expertise
34. Methods for considering the evidence were consistent and transparent, such as through the use of a framework
35. Evidence summary is made available to panel members
36. The prepared evidence summary is transparent and usable for discussion (e.g. knowing where research evidence came from)
37. The quality of the evidence that is used to support the guideline recommendations
38. How evidence is considered and balanced with panel members' input and expert experience
39. The method or process that is used for decision-making in the absence of evidence, or with insufficient evidence
40. The method or process that is used for decision-making with low quality evidence
41. Appropriate consideration is given to all relevant types of evidence
42. Panel members able to provide input and contribute through own expertise and experience
43. How patients' views, perspectives, values, preferences are considered
Formulating the Recommendations
44. The method for formulating the recommendations, such as the use of a framework
45. Transparency of judgements made and providing underlying assumptions and extent of agreement in formulating recommendations
46. Considering setting-specific healthcare factors in formulating the guideline recommendations
47. Considering individual patients' needs and goals when formulating the recommendations
48. Considering the acceptability of the recommendations by end users
49. Considering policy implications and how recommendations are formulated for politically contentious topics
50. Considering the potential of recommendations to impact system change
51. The approach used for wording the recommendation statements
52. Agreement by all panel members on the final recommendations
53. Sufficient explanation of the formulated recommendations to all panel members
54. Transparency of the process from going from the panel's recommendation to the final recommendation that appears in the guideline report
55. No changes being to the recommendations after the panel meeting or when agreement was reached

Consensus
56. The consensus method used by the panel is appropriate, allowing for consensus with diversity of views and not disguising disagreement
57. The panel's ability to reach consensus
58. There is awareness of potential compliance that may lead to spurious consensus
Group Composition
59. The structure of the guideline development group (e.g. may involve a steering committee for logistical and administrative support, patient representatives, internal and external stakeholder, etc.)
60. Diversity in membership and adequate representation of backgrounds and specialties in the panel composition
61. The levels and balance of expertise and methodological support in the panel composition
62. Having patient representatives on the panel
63. Group size is less than 20 members
Group Roles
64. Group members' roles, responsibilities, and tasks are made clear
65. The amount of workload and responsibilities for group members
66. Attendance of all members in the panel meeting(s) (e.g. essential expertise not missing due to panel members' absence)
67. Appropriate involvement of group members throughout the guideline development process
68. Group members adhering to assigned roles and rules
69. Appropriate contribution of group members based on their roles, knowledge and expertise
70. Contributions of all guideline group members are valued
Group Interaction
71. Having environment for open discussion in the panel meeting, with equal opportunity given to all members to contribute to discussion and speak freely
72. Views of all panel members paid attention to and taken into consideration in panel meeting(s)
73. Opportunity given for development of interpersonal relationships and establishment of group norms
74. Mutually respectful relationships fostered between guideline group members
75. Avoiding feeling of need to comply, or abide due to status of some group members and views of authority figure or member with most expertise or confidence
76. Individual group or panel members not dominating the discussion
77. Having opportunity for face-to-face discussion
Group Communication
78. Communication and conduct of meeting(s) is friendly and professional
79. Method of communication with the guideline development group is appropriate and communication is clear
80. Frequency of communication with the guideline development group is appropriate
Incentive
81. There are appropriate incentives for participation in the guideline project
82. Appropriate credit is given for contributions of guideline group members
83. Compensation for involvement in guideline development project
84. There is a perception that involvement in the guideline project will have an impact on health of people
Writing the Guideline
85. How the writing of the guideline is completed

86. Providing input into the draft of the guideline
87. Planning and conducting peer review of the guideline
88. Sufficient time to review the written guideline
Implementation and Dissemination Planning
89. Identification and discussion of research gaps and needs for future research
90. Planning for the dissemination of the guideline
91. Planning for the implementation of the guideline and considering barriers
92. Planning for the assessment of the impact of the guideline
93. There is discussion and agreement about the format(s) of the guideline (e.g. formats for different end users, such as clinician and patient versions, decision on inclusion of care pathways)
Follow-up and Next Steps
94. Evaluation of the guideline development process and feedback from guideline group members
95. Outline for next steps and follow-up clearly communicated to guideline group members

Appendix Table 5: Generalizability analysis for the PANELVIEW tool

Facet	Variance (%)	Interpretation
Panel	0.013 (28)	Variance due to differences between guideline panels
Participants:Panel	0.026 (55)	Variance due to differences between panel members within a panel
Domain	0.002 (4)	Variance due to differences between questionnaire domains
Item:Domain	0.002 (4)	Variance due to differences between items within domains
Panel*Domain	0 (0)	Variance due to differences between domains for any panel
Panel*Item:Domain	0.002 (4)	Variance due to differences between items within domains for any panel
Participants:Panel*Domain	0.001 (2)	Variance due to differences between domains for panel members within panels
Participants:Panel*Item:Domain	0.001 (2)	Variance due to differences between items within domains for panel members within panels
	G	**Interpretation**
Overall generalizability coefficient:	0.35	Overall test reliability to differentiate between panel processes

Appendix Table 5: The generalizability analysis was used to determine the extent to which specific variables (i.e. facets) contribute to the PANELVIEW overall scores. This is represented by the proportion of variance accounted for by each facet. Panel members within a specific panel accounted for the largest proportion of the difference in PANELVIEW scores (55%), while differences in scores between panels accounted for the second largest proportion (28%). PANELVIEW survey domains and items within the domains accounted for a small but non-negligible (4%) proportion of the difference in scores. The overall generalizability coefficient represents the extent to which the PANELVIEW scores can differentiate between panel processes (i.e. those viewed overall as appropriate and satisfactory versus those that are not).

* refers to interaction terms, : refers to nesting of facets within one another (e.g. participants within a guideline panel)

Chapter 4: Appendix 2 - Figures

Appendix Figure 1: Search strategies

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present **Search Strategy: search terms [number of results]**
1 (guideline* adj4 develop*).ti,ab. (21336)
2 (guideline* adj4 process*).ti,ab. (2472)
3 program development/ (28171)
4 guidelines as topic/ or practice guidelines as topic/ (151293)
5 4 and (1 or 2 or 3) (8570)
6 (satisf* or impression* or challenge* or perception* or barrier*).ti,ab. (1396086)
7 attitude*.mp. (407542)
8 6 or 7 (1703467)
9 (participant* or expert* or panel*).ti,ab. (1000160)
10 5 and 8 and 9 (441)
Database: Embase 1974 to Present **Search Strategy: search terms [number of results]**
1 (satisf* or impression* or challenge* or perception* or barrier*).ti,ab. (1762953)
2 attitude*.mp. (499561)
3 1 or 2 (2145587)
4 (participant* or expert* or panel*).ti,ab. (1364650)
5 exp practice guideline/ (523532)
6 (guideline* adj4 develop*).ti,ab. (30241)
7 (guideline* adj4 process*).ti,ab. (3539)
8 6 or 7 (32482)
9 3 and 4 and 5 and 8 (825)

Appendix Box 1: Literature review methods

We included for data abstraction:

- Qualitative or quantitative studies describing evaluation of the guideline development process
- Qualitative or quantitative studies involving interviews or surveys of panelists on their guideline participation experience

Titles and abstracts and full texts of the identified studies were screened independently in duplicate (WW and TB) for inclusion for data abstraction, with disagreements resolved by a third reviewer. We also screened reference lists of included studies.

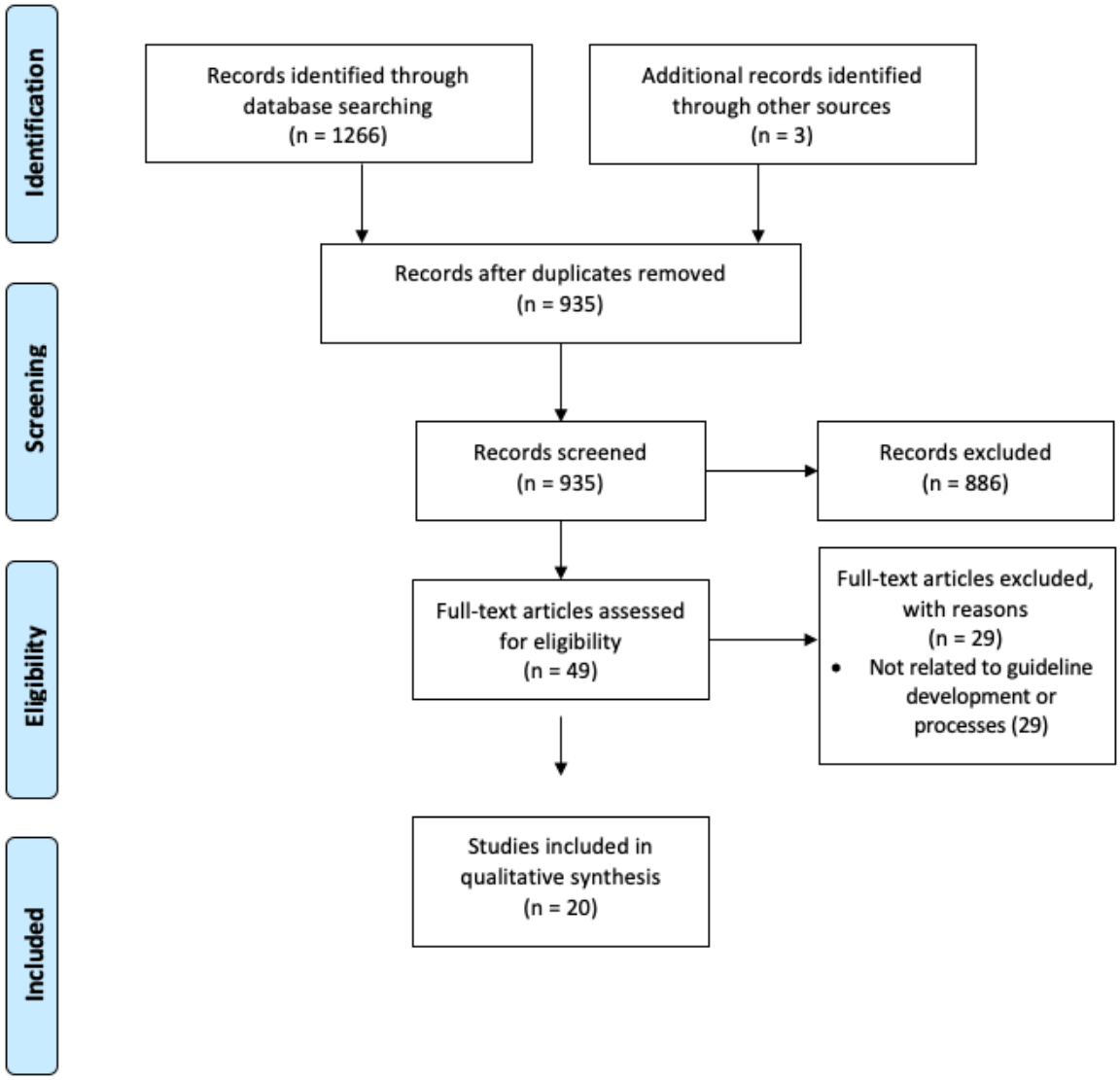
Appendix Figure 2: Item generation survey questions

Panel members participating in guideline panel meetings were approached to provide their feedback about the process they participated in after the meetings adjourned. They were asked to evaluate the process they participated in by responding to the following questions with free-text comments:

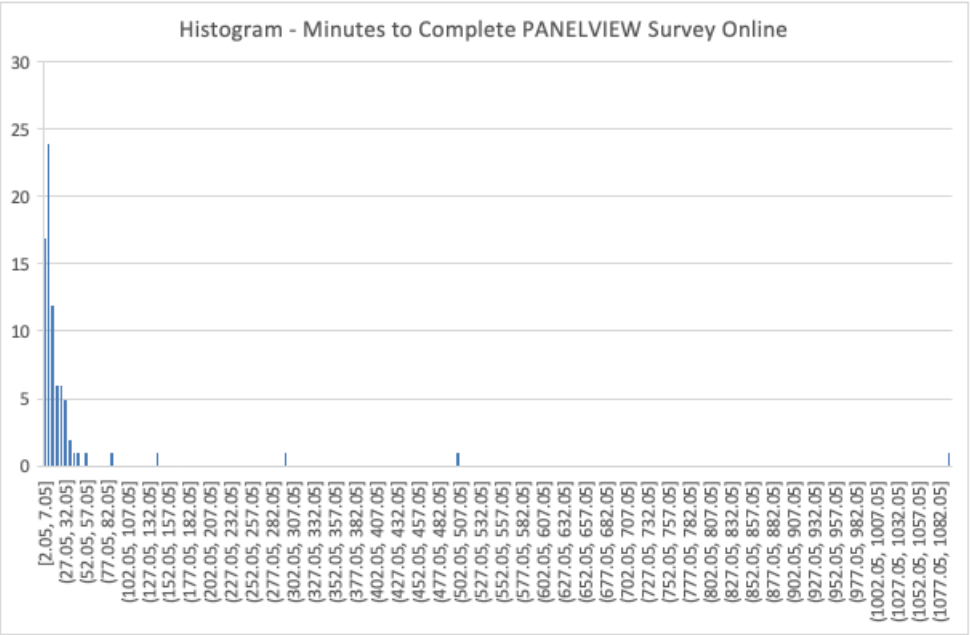
Survey Questions:

1. What were the steps taken during the meeting that helped along the way and made you satisfied with the process?
2. What were the steps taken prior to the meeting throughout the guideline development process that helped along the way and made you satisfied with the process?
3. Were there any issues that made you dissatisfied with the meeting or the overall guideline development process?
4. Please provide your overall impressions of today's meeting and the entire guideline development process. Is there anything else that was done well or wasn't done well? What do you think are the most important parts of the guideline development process that ensure guideline panel members are satisfied? Please specify.
5. Given what you have covered above, what would you identify as the most important steps of the guideline development process that ensure guideline panel members are satisfied?
6. Is there anything else specific to the guideline development process and panel members' satisfaction or any other aspects you would like to mention?

Appendix Figure 3: PRISMA flow diagram for item generation systematic review



Appendix Figure 4: Time to complete the PANELVIEW survey online



Appendix Figure 4: The median time for 80 respondents to complete the PANELVIEW survey was 12 minutes. Removing 12 outliers with a recorded completion time of 30 minutes or longer, who presumably took a break while completing the questionnaire, the median time to complete the survey was 10 minutes and the mean time was 12 minutes.

Appendix Figure 5: Generalizability analysis model

Abbreviations: group (g), participants (p), domain (d), item (i)

ANOVA TABLE

Effect	df	T	SS	MS	VC
g	7	112.98131	112.98131	16.14019	0.01267
p:g	86	1009.17561	896.19430	10.42086	0.28621
d	14	139.78726	139.78725	9.98480	0.01869
i:d	19	249.58278	109.79553	5.77871	0.05064
gd	98	378.62434	125.85578	1.28424	0.00392
gi:d	133	618.22949	129.80962	0.97601	0.04870
pd:g	1204	2016.78894	741.97030	0.61625	0.09356
pi:gd	1634	2924.85598	668.46189	0.40910	0.40910
Mean		0.00000			
Total	3195		2924.85598		

The calculated grand mean = 6.2560
This value has been subtracted from the actual scores for the calculations.

==========

Facets

'g'	Differentiation
'p'	Random
'd'	Fixed
'i'	Fixed

Pattern	Var. Comp.	Levels	Signature	Rule
g	0.013	1.00	d	tau only
p:g	0.026	10.98	dr	Delta and delta
d	0.002	11.33	f	does not contribute
i:d	0.002	25.72	f	does not contribute
g|d	0.000	11.33	df	tau only
g|i:d	0.002	25.72	df	tau only
p:g|d	0.001	10.98 * 11.33	dfr	Delta and delta
p:g|i:d	0.001	10.98 * 25.72	dfr	Delta and delta

RESULTS:
s2(T) = 0.015
s2(D) = 0.028
s2(d) = 0.028
Er2 = 0.345
Phi = 0.345

$$\textit{G-coefficient} = \frac{\sigma_g^2 + \sigma_d^2 + \sigma_{i:d}^2}{\sigma_g^2 + \sigma_d^2 + \sigma_{i:d}^2 + \sigma_{p:g}^2 + \sigma_{p:g|d}^2 + \sigma_{p:g|i:d}^2}$$

CHAPTER 5. DISCUSSION

Summary of Findings

Although guideline development methods have advanced over the past decades, guideline developers continue to encounter challenges and have questions about practical steps they should take to develop trustworthy guidelines. Furthermore, even if followed, guideline developers and users of guidelines, require reassurance that these steps lead to implementing accepted standards that, in turn, ensure that recommendations are trustworthy.

This dissertation addresses some of these knowledge and implementation gaps and describes three research studies that focus on high priority areas for guideline panel engagement. The results of these studies propose new methods and approaches for how panels can define and consider health outcomes, prioritize guideline questions and outcomes, and provide an assessment of the guideline development process they participate in. The findings and implications can be summarized as follows.

The first research study describes a new approach for defining health outcomes using health outcome descriptors (HODs) that help with several areas of confusion and challenges in guideline development and beyond. HODs help with calibrating guideline panel members to an outcome. For example, the term or outcome major bleeding or overdiagnosis of breast cancer, both outcomes considered in guidelines, is interpreted differently by panel members if left explicitly undefined and may therefore lead to different assignment of the magnitude of harms by guideline panels.[1] Creating HODs eliminates or at least reduces this risk by ensuring that panel members use the

definitions as described. HODs help with understanding what outcomes were considered by a panel by making them transparently available, for example, by including them in the guideline and in summary of findings tables. They can help with communication with patients by including them in decision aids. All of this is accomplished with active involvement by guideline panels in the HOD development. This in turn lowers the risk of panel members not agreeing with the outcomes throughout the guideline development. The first study demonstrated one approach for how to develop HODs, the feasibility of the approach, and how they can be used in a guideline development process.

The second research study used the HODs to identify and prioritize health outcomes and, in keeping with the theme of prioritization efforts by guideline panels, tested a novel approach to prioritize guideline questions. Based on findings of two systematic reviews, this was the first study reporting detailed methods for these critical steps.[2,3] Resources and time are limited and guideline developers struggle with identifying the most important questions. Our work provides guidance for how this can be achieved and included a randomized controlled trial comparing different methods to substantiate the approach. One interesting finding was that although panels typically prefer addressing all possible guideline questions, the approach limited, across the 10 guidelines in which we tested it, 25.4% of the questions as high priority. Furthermore, our findings generally suggest that an overall importance rating is sufficient for most scenarios. The overall importance rating was associated with all of six specific questions

that drive prioritization of questions. However, asking detailed questions may help understand which aspects drive the prioritization of questions, e.g. new evidence or cost associated with interventions.

Finally, the third study addressed another aspect of panel involvement by developing an instrument, PANELVIEW, that panel members and others involved in a guideline can use to evaluate if methods described in standards for trustworthy guidelines were appropriately followed. This instrument is entirely novel in that it uses an internal evaluation process as opposed to the approach used by existing instruments that focus on guideline level evaluation based on what is reported rather than what is done, such as the AGREE instrument.[4] It allows for identification of issues in the guideline development process and methods by those who participate in it or directly observe it, and will serve to inform evaluation and quality improvement of guideline programs.

Strengths

The main strengths of the work are that we conducted this research within real guideline development projects and working panels. Therefore, the applicability of our findings should be high, at least for well working guideline panels. Furthermore, our and the advisory group's extensive experience helped focusing on topics that mattered and were informed by prior work. The rigorous development of each of the research projects is another strength.

Limitations and Future Research

A remaining issue with regards to HODs is that once developed they require vetting by other stakeholders including trialists, those working in quality improvement, systematic review authors, health technology assessors and in other areas.[5] A challenge related to the development of HODs and completion of outcome utility rating was that some participants misinterpreted the two ends of the visual analogue scale and, thus, provided invalid ratings. A limitation of our work on question prioritization is that we did not explore whether the overall importance rating or the criteria ratings on the individual questions predict the final approval by the guideline panel for inclusion. Utility ratings of outcomes by guideline panels will help informing future guideline efforts, but we will have to evaluate if they correlate with utility ratings obtained in other contexts (e.g. from patients). Comparisons between ratings by guideline panels with those of external stakeholder groups should be undertaken and an integration with efforts such as Core Outcome Measures in Effectiveness Trials (COMET, http://www.comet-initiative.org/) should be undertaken by us or others. For the PANELVIEW instrument development, we have only a limited number of panels with members that provided overall low ratings across the instrument domains and items. This is probably related to the fact that we recruited panels from well documented guideline efforts. In future work, the PANELVIEW instrument will need to be administered with additional, diverse panels from various guideline organizations, to better assess its ability to discriminate guideline processes at the panel level.

In follow up work to each of the three topics of our research, comparative studies will allow further assessment and impact of the proposed methods and approaches. Further validation of HODs could be assessed in RCTs on the impact of using and not using HODs with respect to judgements of importance of outcomes, utility of outcomes and decision-making. RCTs evaluating the approaches for question and outcome prioritization with panels would assess if they lead to different outcomes or questions being considered high priority and critical. Comparative studies with the PANELVIEW instrument would allow comparing PANELVIEW assessments of guideline panelists with that of other group members (e.g. non-voting observers).

Conclusion

In conclusion, guideline development is complex and multidisciplinary, requiring coordination and execution of a number of steps and involvement of multiple stakeholder groups. Indeed, it has become more complex with advancement of new methods and standards, but this complexity is justified as guidelines are a major driver of practice and selection of options to deliver the best health care. Therefore, it is obvious and critical that the most trustworthy approaches are used. To achieve trustworthiness, guideline panel member education, calibration to the task and engagement are important to ensure that guidelines are focused on the population they are attempting to serve. Guideline panels play a key role and it is important to ensure they are able to adequately complete the steps of developing trustworthy guideline and reach our standards for best health care. The research summarized in this thesis

addresses three priority areas that deliver methods and tools for improving rigour in guideline development, transparency, and inform quality improvement of future guidelines.

References

1. Baldeh T, Saz-Parkinson Z, Muti P, et al. Development and use of health outcome descriptors: a guideline development case study. *Health Qual Life Outcomes.* 2020;18(1):167.
2. El-Harakeh A, Lotfi T, Ahmad A, et al. The implementation of prioritization exercises in the development and update of health practice guidelines: A scoping review. *PLoS One.* 2020;15(3):e0229249.
3. El-Harakeh A, Morsi RZ, Fadlallah R, Bou-Karroum L, Lotfi T, Akl EA. Prioritization approaches in the development of health practice guidelines: a systematic review. *BMC Health Services Research.* 2019;19(1):692.
4. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2010;182(18):E839-842.
5. Schunemann HJ, Tunis S, Nieuwlaat R, Wiercioch W, Baldeh T. The SOLAR (Standardized Outcomes Linking Across StakeholdeRs) system and hub and spokes model for direct core outcome measures in health care and its relation to GRADE. *J Clin Epidemiol.* 2020.