PRIFYSGOL BANGOR / BANGOR UNIVERSITY

ROBVALU: A tool for assessing risk of bias in studies about peoples' values, utilities, or the importance of health outcomes

Karam, Samer G.; Zang, Yuan; Pardo-Hernandez, Hector; Siebert, Uwe; Koopman, Laura; Noyes, Jane; Tarride, Jean-Eric; Stevens, Adrienne; Welch, Vivian; Parkinson, Suleika Saz; Ens, Brendalynn; Devji, Tahira; Xie, Feng; Hazlewood, Glen; Mbuagbaw, Lawrence; Coello, Pablo Alonso; Brozek, Jan L.; Schünemann, Holger J.

BMJ

DOI:
10.1136/bmj-2024-079890

Published: 12/06/2024

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):
Karam, S. G., Zang, Y., Pardo-Hernandez, H., Siebert, U., Koopman, L., Noyes, J., Tarride, J.-E., Stevens, A., Welch, V., Parkinson, S. S., Ens, B., Devji, T., Xie, F., Hazlewood, G., Mbuagbaw, L., Coello, P. A., Brozek, J. L., & Schünemann, H. J. (2024). ROBVALU: A tool for assessing risk of bias in studies about peoples' values, utilities, or the importance of health outcomes. *BMJ*, *385*, Article e079890. https://doi.org/10.1136/bmj-2024-079890

Hawliau Cyffredinol / General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

19. Jun. 2024

ROBVALU: A tool for assessing risk of bias in studies about peoples' values, utilities, or the importance of health outcomes

Samer G. Karam[1,2], Yuan Zhang[1,2], Hector Pardo-Hernandez[3,4], Uwe Siebert[5,6,7], Laura Koopman[8], Jane Noyes[9], Jean-Eric Tarride[1,10,11], Adrienne Stevens[12], Vivian Welch[13], Zuleika Saz Parkinson[14], Brendalynn Ens[15], Tahira Devji[16], Feng Xie[1,10], Glen Hazlewood[17,18], Lawrence Mbuagbaw[1,19,20,21,22,23], Pablo Alonso Coello[3,4], Jan L. Brozek[1,2,24], Holger J. Schünemann[24,25]

1) Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
2) Michael G. DeGroote Cochrane Canada & McMaster GRADE Centres, McMaster University, Hamilton, Ontario, Canada
3) Iberoamerican Cochrane Centre, Sant Pau Biomedical Research Institute (IIB Sant Pau), Barcelona, Spain
4) CIBER Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, Spain
5) Department of Public Health, Health Services Research and Health Technology Assessment, Institute of Public Health, Medical Decision Making and Health Technology Assessment, UMIT TIROL – University for Health Sciences and Technology, Hall i.T., Austria
6) Center for Health Decision Science and Departments of Epidemiology and Health Policy & Management, Harvard T.H. Chan School of Public Health, Boston, MA, USA
7) Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
8) Department of Specialist Medical Care, National Health Care Institute, Diemen, The Netherlands.
9) School of Social Science, Medical and Health Sciences, Bangor University, Wales, UK
10) Center for Health Economics and Policy Analysis (CHEPA), McMaster University Faculty of Health Sciences, Hamilton, Ontario, Canada
11) Programs for Assessment of Technologies in Health, St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada
12) Centre for Immunization Readiness, Public Health Agency of Canada, Ottawa, Canada
13) Bruyère Research Institute and, School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
14) European Commission, Joint Research Centre (JRC), Via E. Fermi 2749, 21027, Ispra, VA, Italy
15) Canadian Agency for Drugs and Technology in Health, Ottawa, ON, Canada
16) Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.

17) Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
18) Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada
19) Department of Anesthesia, McMaster University, Hamilton, ON, Canada
20) Department of Pediatrics, McMaster University, Hamilton, ON, Canada
21) Biostatistics Unit, Father Sean O'Sullivan Research Centre, St Joseph's Healthcare, Hamilton, ON, Canada
22) Centre for Development of Best Practices in Health (CDBPH), Yaoundé Central Hospital, Yaoundé, Cameroon
23) Division of Epidemiology and Biostatistics, Department of Global Health, Stellenbosch University, Cape Town, South Africa.
24) Clinical Epidemiology and Research Center, Department of Biomedical Sciences, Humanitas University, Milan, Italy
25) Humanitas Research Hospital, Via Rita Levi Montalcini 4, 20090 Pieve Emanuele, Milan, Italy

***Corresponding author**

Prof. Holger Schünemann
Clinical Epidemiology and Research Center (CERC)
Department of Biomedical Sciences
Humanitas University
Via Rita Levi Montalcini 4 – 20090 Pieve Emanuele (Milano) Italy
E-mail: schuneh@mcmaster.ca
Tel: +1 905 525 9140 x 24931
Fax: 1 905 522 9507

Standfirst

People's values are important drivers in health-care decision making. The certainty of an intervention's effect on benefits and harms relies on two factors: the certainty in the measured effect on an outcome in terms of risk reduction and the certainty in its value, utility or importance. The GRADE working group has proposed a set of questions to assess risk of bias (ROB) in a body of evidence from studies addressing how people value outcomes. However, no validated ROB tool in individual value, utility, and importance of outcome studies exists, which is required to evaluate such evidence.

Hence, we developed the ROB in VALues and Utilities (ROBVALU) tool. ROBVALU has good psychometric properties and will be useful when assessing individual studies in measuring values, utilities, or the importance of outcomes. As such, ROBVALU can support health research assessments, where the certainty of input variables determines the certainty in model outputs, for example, in decision-analytic benefit-harm analysis for health guidelines and cost-utility or cost-effectiveness analysis for health policy and reimbursement decision-making.

Summary Box

- The risk of bias (ROB) in VALues and Utilities (ROBVALU) tool serves to assess risk of bias in studies determining values, utilities, or importance of outcomes studies
- The tool covers four separate subdomains through which bias might be introduced
- The individual subdomain judgments inform the studies' overall ROB
- ROBVALU has demonstrated high validity and reliability

Introduction

Healthcare decision-making relies on evidence on the relative effectiveness, safety and cost-effectiveness of an intervention evaluated in appropriate studies [1, 2]. Choosing between different interventions, such as a preventive, diagnostic or treatment strategies, depends on the importance or value people place on specific health states or health outcomes [2]. Values play a major role at different levels of decision making, from the individual to the healthcare system level. In this context, people's values reflect the importance they place on outcomes of interest that result from decisions about using an intervention, e.g., taking a certain test or starting a new treatment regimen [2]. We use the term "people" when talking about value as the term is inclusive to patients, healthcare providers, policy makers, and the general public. Utility instruments are widely used to elicit the absolute value of a health outcome and provide an index measure anchored on a scale with 1 reflecting "perfect health" and 0 reflecting "being dead".[3, 4]. Indeed, various methods are used to establish values, including direct measures of utility, indirect measurements of utility, or qualitative research [2, 5]. The visual analogue scale (VAS) is one of the simplest measures to elicit these values. People are asked to rate a health state on a VAS that is then converted to a utility value [6, 7]. While the VAS directly measures the importance of an outcome, concerns exist about how accurate and valid it may be [2]. Other direct measures such as the standard gamble and time-trade-off require people to choose between their current health state and a treatment option that may result in perfect health or in immediate death [4, 8]. Discrete-choice experiments ask people to choose between two or more treatment options, where the choices differ in terms of their attributes, that are defined by the investigators [9]. The relative importance of each attribute is then inferred by analyzing the responses, assuming patients choose the option with the highest value [9]. Indirect methods of measuring

utility values include validated health related quality of life (QoL) instruments, such as the EQ-5D and the Health Utilities Index (HUI) [10]. The EQ-5D requires respondents to answer questions across five domains that are converted to a utility value using validated scoring systems [11, 12].

General application of utility values in research

These utility values allow weighing the benefits and harms of an option and, thus, they also play a cardinal role in health economics and health technology assessments [3, 13]. For instance, in decision analysis they are required to calculate quality adjusted life years (QALY). Confidence in studies that report on values needs to be ascertained for decision-making in guideline recommendations, health technology assessments, or coverage decision [14]. For example, in a systematic review on people with chronic obstructive pulmonary disease, we found that there is moderate certainty that patients value adverse events as important, but on average less important than symptom relief [15]. We also found moderate certainty that exacerbation and hospitalisation due to exacerbation are the outcomes that COPD patients' rate as most important. In another example, a systematic review on patients values on venous thromboembolism (VTE), we found that people with cancer place more importance on a decrease of new or recurrent VTE than on a decrease in major or minor bleeding events [16].

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks, a widely approach used in guidelines, health technology Assessment and other decisions, require judgments about the certainty in how much people value the main outcomes: "Is there important uncertainty about … how much people value the main outcomes?"[17, 18]. One of the key determinants of certainty is internal validity,

that is, how well individual studies were designed and conducted, i.e., internal validity which GRADE and Cochrane label as the risk of bias (ROB) domain.

Risk of bias

Similar to other study designs, threats to internal validity arising from the study design, conduct, analysis, and reporting of the study introduce ROB in research on utility values [2]. Poor study quality could result in indirectness which encompasses applicability and external validity, often as a result of PICO elements. Another quality issue is low sample size or no sample size calculation which may result in imprecision. ROB assessment tools are developed to assess biases that result in threats of internal validity and would not measure indirectness and precision. Quality assessment tools and reporting checklists often address all factors of a studies qualities and safeguards, this is different form a ROB assessment tool that aims to present a ROB judgment for a study. One key factor that may introduce bias in values studies is the measurement instrument used to measure utilities of the people in the study. Bias means that a value people place on an outcome in a research study, e.g., a value of 0.5 for stroke, would be systematically different from the true value that people would place on that outcome. That is, the true, unbiased value may be 0.3 and, thus, using biased estimates would provide wrong answers in the modeling and the health decision-making context.

ROB assessment tools exist for many study designs including the Cochrane Risk of Bias 2 (RoB 2) for randomised trials [19], ROBINS-I for non-randomised studies of the effects of interventions (NRSI) [20] and ROBINS-E for studies about exposures [21, 22]. Critical appraisal tools to assess the quality of a study, such as the Newcastle-Ottawa scale and the JBI critical appraisal tool for cross-sectional studies, are also study design specific[23, 24]. These tools are regularly used by researchers to assess the quality of individual studies or to assess ROB,

however, they were not developed for utility values studies. These checklists invariably include questions that are study design specific that would not always be appropriate to address in studies about peoples values (e.g., “Were there deviations from the intended intervention that arose because of the trial context?” or “Was the exposure measured in a valid and reliable way?”). A major concern with utility values studies which is not adequately addressed by any commonly used ROB tool is the method used to elicit peoples values. The measurement instrument needs to be valid and reliable, administered appropriately, valid health outcomes used, and proper understanding of the instrument explored. No validated tool for the nuanced assessment of the ROB in individual studies measuring utility values is available [9, 20, 25-27].

Objective

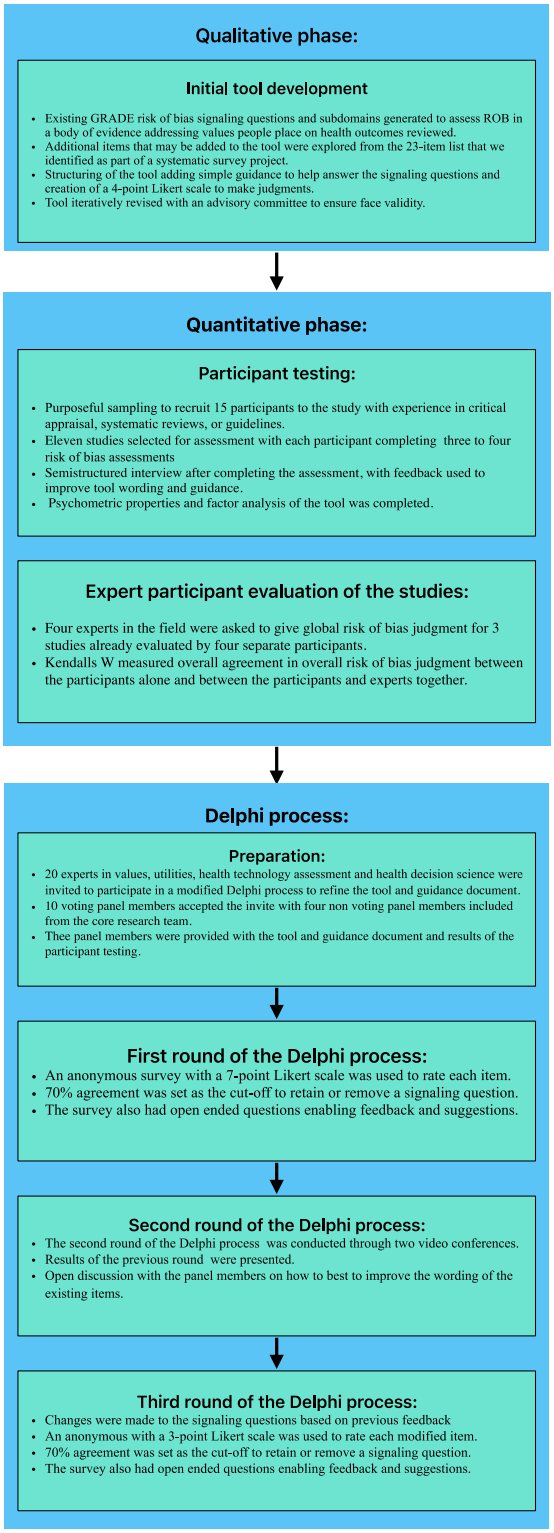
To properly implement evidence-based decision-making and formulate evidence-based recommendations in clinical or public health guidelines, it is crucial to evaluate ROB in values, utilities, or importance of outcome studies. However, due to the absence of specialized and validated ROB assessment tool this is rarely done. Thus, our goal was to develop, validate, and describe a pragmatic ROB tool for studies measuring the value people place on health outcomes with appropriate guidance to apply it correctly.

Development of the ROBVALU tool and guidance

We followed a sequential mixed-methods approach starting with a qualitative approach to develop ROBVALU and related guidance document (Supplement S1) [28], followed by a quantitative phase to assess the psychometric properties of the tool (Figure 1). In the qualitative phase, we began by considering the ROB signaling questions (Table A1 in the Appendix) and subdomains that we had carefully developed for GRADE guidance to assess ROB about values

across studies in a body of evidence [2]. For that GRADE guidance, we iteratively developed the subdomains and signaling questions starting with a 23-item list that we identified as part of a systematic survey project [27]. The core research group reviewed the 23-item list to identify any missing item that may be relevant for the single study ROBVALU tool, after thorough discussions within the group a decision was made not to add any new items or subdomains to avoid complexity, improving applicability, feasibility, and adoption of the tool. We first structured a preliminary version of the tool and added simple considerations to help answer the signaling questions. These signaling questions were categorized into four subdomains: Selection of participants into the study, completeness of data, measurement instrument, and data analysis. We used a 4-point Likert-type scale (yes, probably yes, probably no, no) to judge the individual items, this was done to avoid a neutral option of a 5-point Likert scale when studies lack sufficient information to make a proper judgment. In each subdomain the tool asked for how important and how serious the risk of bias issue is. The core research group iteratively revised the ROB tool and the accompanying guidance document. An advisory group of experts provided feedback and suggested appropriate changes to establish face and content validity (Supplement S2).

Figure 1. Tool development process



Participant testing

We used purposeful sampling to recruit 15 participants with experience in critical appraisal, systematic reviews, or guidelines for user testing and semi-structured interviews (supplement S3). The participants had a broad level of expertise and included master level students to senior researchers with experience in health research ranging from 6 months to 30 years (Table A2 in the Appendix). All users received the ROBVALU tool and the accompanying guidance document (Supplement S1). We instructed the participants to complete three to four assessments and every sample study was assessed by four users independently, 11 studies in total were assessed (Table A3 in the Appendix). Based on feedback received in the semi-structured interview after user testing, we iteratively revised and improved the guidance document throughout the project with a focus on the wordings, spelling, and grammatical structure of the guidance document. The ROBVALU tool demonstrated good psychometric properties with an overall intraclass correlation coefficient of 0.87 and the four subdomains showed good to excellent reliability ranging from 0.80 to 0.91 (Table 1 and Supplement S4). We also calculated the inter-rater reliability of the global ROB judgment using the ROBVALU tool using Kendall's \ W that showed substantial agreement of 0.62 (Supplement S4). We invited four expert participants in the field to provide a global judgment for ROB without using the ROBVALU, with each expert rating three to four studies. When we added the expert participant responses of the global ROB judgment the Kendall's W dropped to 0.45 showing moderate agreement (Supplement S4). However, only four global judgment responses were more than one level of seriousness higher or lower than the expert participant judgment (Table A4).

Table 1. Reliability of ROBVALU

Subdomain	Cronbach's Alpha
Selection of participants	.87 (95%CI: 0.79-0.93)
Completeness of data	.90 (95%CI: 0.84-0.94)
Measurement instrument	.80 (95%CI: 0.69-0.88)
Data analysis	.91 (95%CI: 0.86-0.95)
Total	.86 (95%CI: 0.78-0.91)

Modified Delphi process

Finally, following our protocol, we used purposeful sampling to invite 20 experts in values, utilities, health technology assessment and health decision science to participate in a modified Delphi process for final refinement of the tool (Supplement S5, Figure S8) [29-31]. We used our extensive network of global colleagues working in the field of study to identify and invite the expert panel. Ten voting members accepted the invite to participate in the Delphi panel, and four members of the working group participated as non-voting members. We shared the ROBVALU tool draft, guidance document, and the results of our participant testing with the panel members. The first round of the Delphi process involved an anonymous survey to determine the signaling questions to be included. The second round took place via recorded video conferences with the aim of identifying common themes and reaching consensus on simplifying and harmonising language across the tool. The third and final round of the Delphi process included an anonymous survey for final consensus on the wording of the signaling questions and the proposed methods for providing a global ROB judgment. We used google forms to prepare the surveys, and in the first survey we used a 7-point Likert scale (strongly agree, agree, somewhat agree, neutral,

somewhat agree, disagree, and strongly disagree) to rate each item, with 70% agreement set as the cut-off to retain or remove a signaling question. In the final survey we used a 3-point scale (agree, neutral, and disagree) with 70% agreement set as cut-off to retain the signaling question. In the first round of the Delphi process, we had 100% response rate resulting in 80% to 100% consensus to retain all signaling questions. We also collected feedback from open ended questions for suggested edits for the signaling questions (Supplement S6). In the second round of the Delphi process, we presented the ROBVALU tool, the psychometric properties, the exploratory factor analysis, and the results of the first round of the Delphi to the panel members. After deliberating on the tool's properties, agreement was reached to edit some signaling questions to simplify the language or to harmonize the language across the tool. This resulted in minor changes only. We also discussed how to make a final judgment for ROB for a study. We had 100% response rate in the third and final round of the Delphi process resulting in 80% to 100% consensus on the tools signaling questions, including the ones with minor adjustments to the wording. We also established consensus of >70% that the overall ROB judgment should match the most severe ROB judgment on an item unless the appraisers can provide justifications to rate the overall ROB lower (e.g., many concerns on many items) or higher (concern seems to not be influencing overall ROB importantly). For example, if multiple subdomains were rated as very serious, the final judgment could be rated as extremely serious (Supplement S7).

Risk of bias subdomains

ROBVALU included seven key signaling questions across four subdomains selection of participants into the study, completeness of data, measurement instrument, and data analysis (Table 2).

272

273 **Table 2. Risk of bias subdomains and considerations in ROBVALU**

Risk of bias subdomains	Signaling questions	Rationale/ Example
Selection of participants	***Was an appropriate study sample selected from the study's sampling frame?*** *(Consider: what is the sampling strategy? i.e., random sample or consecutive sample, etc. Is there a subset of the population that is more or less likely to be reached with this sampling strategy?)* ● *Yes* ● *Probably yes* ● *Probably no* ● *No*	Reviewers should determine whether the sampling strategy was conducted in a manner to minimize the risk of selection bias. In a comparison study, selection bias refers to systematic differences between baseline characteristics of the groups that are compared. Here, for risk of bias, we only refer to bias internal to the study, rather than inadequate generalizability (applicability or "directness"); that is, selection bias that could happen when the achieved sample is deviated from the intended sample (as described in the protocol or the methods section of the study), rather than from the population we intend to extrapolate the conclusion to (i.e., the target population of the research question). We need to assess to what extent the achieved sample is similar to the intended sample. The sampling strategy is a critical component since it will influence the results through the population the researcher's had studied. For example, for a cross-sectional study, a stratified random sampling strategy would minimize the risk, while a convenience sample would probably be a biased sample for the study population.
Completeness of data	***Was the attrition rate sufficiently low to minimize the risk of bias?*** *(To consider: what was the response rate? If follow-ups were planned and used, what was the attrition rate during the follow up? Were the participants responded*	In addition to sampling strategy, in surveys, response rate also influences the representativeness of the achieved sample. The higher the response rate the less likely risk of bias is a concern. Response could be influenced by various factors, including study design, study purposes, sampling strategy, and survey administration. There is no single rule for an "inadequate" response rate though; if the judgment is not an acceptable response rate, provide justification. For longitudinal studies with follow-ups planned and used, the attrition rate such as drop-outs, loss to follow up and exclusions could be another source of concern

	systematically different from those not?) ● *Yes* ● *Probably yes* ● *Probably no* ● *No*	
Measurement instrument	***Was the instrument used to measure patient values and preferences in a valid and reliable manner?*** *(Consider: what was the measurement instrument selected? does the instrument have well-constructed validity and reliability? Or is this instrument widely accepted in this area to have adequate reliability and validity?). (Translation and culturally adapted in guidance)* ● Yes ● Probably yes ● Probably no ● No	Measurement instrument refers to direct measures of utility (e.g., standard gamble and time trade-off, conjoint analysis with discrete choice experiments) and indirect measurement instruments of utility such as EQ-5D. A variety of measurement instruments could be chosen, including those providing utility measurements (standard gamble, time trade off, visual analogue scale, etc.), willingness to pay, discrete choice, or other structured scales. For a specific study, the validity and reliability of the instrument may not always have been determined. In these cases, to be considered a reliable and valid instrument, either the researchers provide the validity and reliability information in the study being evaluated, or the measurement instruments are widely accepted as both reliable and valid.
	Was the instrument administered in the intended way? ● Yes ● Probably yes ● Probably no ● No	Faulty measurements could be a source of bias, either due to inherent shortcomings in a measurement tool or via administration error. For a specific study, the researchers should demonstrate the measurement tools were administered correctly or in a manner conforming to their rationale to minimize the risk of introducing bias. If applicable, tools should be administered in a consistent manner across different subpopulations.

	Was a valid representation of the outcome (health state) utilized? • Yes • Probably yes • Probably no • No	The description of health states is another possible source of bias. High quality description provides participants with best available evidence, while wrong or insufficient information based on low quality evidence may mislead participants and bias the measurement. High quality description consists of the experience, probability, duration, and consequences of a health state and should be presented in an understandable format.
	Did the researchers check for understanding of the instrument? • The investigator tested the understanding, and understanding was adequate; • The investigators did not formally test the understanding, but there was evidence suggesting adequate understanding • The investigators did not formally test the understanding; but there was evidence suggesting inadequate understanding.	If the participants have problems to understanding the techniques, the results they provide are likely to be misleading. There is a gradient in the understanding of measurement techniques. Depending on whether the understanding is checked formally, and whether the understanding is adequate.

	● The investigator tested the understanding, but understanding was inadequate.	
Data Analysis	***Were the results analyzed appropriately to avoid influence of bias and confounding?*** *(Consider whether the adjustment, stratification, strategy to deal with missing data and model selection, if any, was appropriate)* ● Yes ● Probably yes ● Probably no ● No	The appropriateness of data analysis would include the strategy to deal with missing data and/or excluded cases from analysis. If confounding factors or other influential factors exist, statistical techniques such as stratification or regression analyses for adjustment of measured confounding factors may be taken when appropriate. Often, in an outcome valuation study, no adjustment is made, and the results are reported in different subgroups. Furthermore, the appropriateness of model selection (if any) or analysis strategy should be checked.

Selection of participants into the study

Precise research questions include a clear definition of the target population. The study population of any empirical study must be representative for this target population, and is therefore, a critical component since bias in the selection will lead to biased estimates of the values people place on outcomes in the target population [2]. When assessing selection bias, users should consider the study's sampling strategy, in particular if the achieved sample population deviates from the intended sample population [2], as this may lead to biased estimates for the study's population of interest due to threats to internal validity. If the achieved sample

population does not deviate from the intended sample population, but it differs from the population one intends to extrapolate the results to, it will result in lack of generalizability. We refer to it as indirectness which encompasses applicability and external validity. The ROBVALU tool is not intended to address indirectness, a different domain in assessing the certainty of a body of evidence according to GRADE, but we are developing a tool that is specific to indirectness separately.

Completeness of data

When judging completeness of data, reviewers need to consider the response rate of the study population, the attrition rate if follow-up was involved, and the differential responders compared to non-responders [2]. High response rates and/or low proportion of loss to follow-up are clearly preferable, and a high proportion of nonresponse or dropout could be problematic [2]. Participants providing responses may very plausibly differ from those who do not, and to the extent this is the case, results coming only from those who responded or completed follow-up may be misleading [2].

Measurement instrument

It is important to use reliable and valid instruments to measure the relative importance of outcomes in values, preferences, and utility studies [2]. Using unreliable or poorly validated instruments can result in biased measurements of the outcome. Similarly, utility values for specific health-states based on instruments not sufficiently validated that are used as input parameters for decision-analytic models can result in biased estimates, such as quality-adjusted life years (QALYs) derived from state-transition models[32, 33]. Researchers conducting

306 primary empirical studies should provide information regarding the measurement properties of
307 the instrument they have chosen [2].
308 Researchers should also demonstrate that the instrument has been administered correctly and in a
309 consistent manner across all participants in a study. For example, if the standard gamble is to be
310 administered by an interviewer, self-administration would pose risk of bias as utility estimates
311 could be systematically different. In addition, an optimal representation of the outcome or health
312 state should be presented/described in a way that accurately reflects the attribute the researchers
313 intended to measure. This may include a detailed explanation of how the outcome defines the
314 experience, the probability of the outcome, durations, and possible consequences. Finally, it
315 should be evaluated as to whether participants had a proper understanding of the instrument to
316 complete the tasks.
317
318 ***Data analysis***
319 Studies should explore heterogeneity in values when appropriate and present results for the
320 different subgroups. The data analysis plan and exploration of heterogeneity should be outlined a
321 priori before collection of data. A causal framework that helps delineate health state and outcome
322 interactions with possible confounding factors will help make assumptions explicit. If
323 heterogeneity is found, the evaluator needs to consider whether the adjustment, stratification, or
324 model selection used in the study reporting on values was appropriate [2]. Adjusting for
325 important confounding factors, such as age if it is associated with the intervention and influences
326 the estimated values, or reporting values in stratified manner, reduces biased estimates of the
327 value placed on an outcome. In addition, self-inflicted biases, including selection bias or

328 immortal time bias should be controlled for appropriately using modern causal inference
329 methods (e.g., target trial emulation or g-methods for time-varying confounding)[34].
330
331 **ROBVALU tool application**
332 The assessment of ROB in studies evaluating the value people place on outcomes involves the
333 following steps:
334 1) specify the research or review question;
335 2) specify the outcome being assessed;
336 3) identify the sampling frame, the response rate and/or attrition rate, the measurement
337 instrument used, and the data analysis plan;
338 4) answer the signaling questions of the four subdomains;
339 5) make a judgment if there are important risk of bias concerns in the four subdomains;
340 6) formulate a risk of bias judgment for the four subdomains;
341 7) formulate an overall risk of bias judgment for the study outcome being assessed.
342 The ROBVALU tool (Table 3) provides users with space to record vital information of the study
343 being assessed, and signaling questions to all four subdomains that must be addressed. We
344 validated a 4-point Likert-type scale (yes, probably yes, probably no, no) to respond to the
345 individual signaling questions (items). When rating individual signaling questions, we suggest
346 following the flowchart (Figure 2) for consistent answers between raters. In each subdomain the
347 tool asks to specify how important the ROB issue is on a 4-point Likert-type scale (yes, probably
348 yes, probably no, no), and then how serious the overall ROB issue is on a 4-point Likert-type
349 scale (not serious, serious, very serious, extremely serious). Responses to the signaling questions
350 should provide the basis for the subdomain level judgment, of how important and how serious

the ROB issues are in the study. Raters should provide a rationale for the response as free text, to justify their judgments. We suggest that the final judgment for each subdomain inversely correlates with the signaling question judgment. For example, in the measurement instrument subdomain, if the answer to “Was the instrument administered in the intended way?” was “No”, then the answer to “Are there important risk of bias issues concerning the measurement instruments?” should be “Yes”. If raters believe that the lowest signaling question judgment does not reflect the overall subdomain judgment, they may choose not to deem the results of the study at risk of bias for that subdomain, but they are asked to provide explanations for why they would not do this.

362 **Table 3. Risk of bias assessment tool: ROBVALU**

RISK OF BIAS ASSESSMENT TOOL(ROBVALU)
Rating the Risk of Bias of Research Evidence in Studies on Values, Utilities, or the Importance of Outcomes

Specify the study question

Selections of participants	
Completeness of data	
Measurement instrument	
Data analysis	

Specify which outcome is being assessed

We suggest that the final judgment for each subdomain inversely correlates with the lowest judgment of the signaling question. If you believe that the lowest subdomain judgment does not reflect the overall subdomain judgment, they may choose not to deem the results of the study at risk of bias, but they should provide explanations for why they would not do this.

SELECTION OF PARTICIPANTS INTO THE STUDY

Was an appropriate study sample selected from the sampling frame? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

Consider the study's sampling strategy, is it a random sample or consecutive sample? Is there a subset of the population that is more or less likely to be reached with this sampling strategy?

Rationale:

Are there important risk of bias issues concerning selection of participants into the study? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

How serious are the risk of bias issues concerning selection of participants into the study? ☐ Extremely serious ☐ Very serious ☐ Serious ☐ Not serious

COMPLETENESS OF DATA

Was the attrition sufficiently low to minimize the risk of bias? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

Consider the response rate; if follow-up was involved, also the attrition rate; and the characteristics of the participants who responded and those who did not.

Rationale:

Are there important risk of bias issues concerning completeness of data?	☐ Yes	☐ Probably yes	☐ Probably no	☐ No
How serious are the risk of bias issues concerning completeness of data?	☐ Extremely serious	☐ Very serious	☐ Serious	☐ Not serious

MEASUREMENT INSTRUMENT

Was the instrument used to measure patient values and preferences in a valid and reliable manner? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

Consider if the instrument chosen is familiar to assessors and is widely accepted to be reliable and valid, Also, consider whether the authors provide information regarding the measurement properties of the instrument chosen. Consider if the tool used is a validated translation.

Rationale:

Was the instrument administered in the intended way? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

Consider whether the instrument was administered correctly, and in a consistent manner across participants and subpopulations.

Rationale:

Was a valid representation of the outcome (health state) utilized? ☐ Yes ☐ Probably yes ☐ Probably no ☐ No

Optimal representation of the outcome includes a detailed explanation of how the outcome that defines the experience, probability, duration, and consequences was developed. This question only applies when the participants are asked to indicate the importance they would like to place on a set of hypotheticals or described outcomes, rather than their own health.

Rationale:

Did the researchers check the understanding of the instrument?

- ☐ The investigators tested the understanding, and understanding was adequate
- ☐ The investigators did not formally test the understanding, but there was evidence suggesting adequate understanding
- ☐ The investigator tested the understanding, and understanding was inadequate
- ☐ The investigators did not formally test the understanding, but there was evidence suggesting inadequate understanding

Consider whether the investigators piloted the study, or if the instrument was simple enough to assume understanding.				
Rationale:				
Are there important risk of bias issues concerning the measurement instruments?	☐ Yes	☐ Probably yes	☐ Probably no	☐ No
How serious are the risk of bias issues concerning measurement instruments?	☐ Extremely serious	☐ Very serious	☐ Serious	☐ Not serious

DATA ANALYSIS				
Were the results analyzed appropriately to avoid influence of bias and confounding?	☐ Yes	☐ Probably yes	☐ Probably no	☐ No
Consider whether the adjustment, stratification, or model selection was appropriate. Was there a priori analysis plan.				
Rationale:				
Are there important risk of bias issues concerning the data analysis?	☐ Yes	☐ Probably yes	☐ Probably no	☐ No
How serious are the risk of bias issues concerning data analysis?	☐ Extremely serious	☐ Very serious	☐ Serious	☐ Not serious

OVERALL RISK OF BIAS FOR THE STUDY	
Not serious	☐
Serious	☐
Very serious	☐
Extremely serious	☐

The global ROB judgment for a study corresponds to the lowest subdomain judgment (Table 4), this is done because any domain level bias will lower our confidence in the study results. If users do not believe that the lowest subdomain judgment reflects the global ROB judgment, they should provide a justification. For example, if a study has a low response rate resulting in very "serious risk of bias" domain judgment and the study results are comparable to better quality

studies, a reviewer may consider that the subdomain judgment does not reflect the global ROB judgment. An illustrative example of a completed assessment is provided in Box 1.

Box1. ROBVALU to assess the risk of bias in values assigned to a exacerbation of chronic obstructive pulmonary disease (COPD) [35]

In assessing the utility value patients with chronic obstructive pulmonary disease (COPD) place on an exacerbation, a study evaluated 65 males and females with COPD at 7 study sites in the United States when they visited an outpatient clinic withing 48 hours of symptom onset [35]. Participants had to be 40 years or older and had to be current or former smokers with a history of at least 10 pack-years. Of 65 subjects, 59 completed the study and 3 were lost to follow up and 3 were ineligible. The utility values were measured using the EQ-5D.

- Selection of participants into the study likely lead to risk of bias: Exacerbations that required hospital admission were considered severe and were excluded from this study and might thus importantly bias. Thus, the population was deemed to be probably not representative of the intended population. A risk of bias assessment using the ROBVALU tool revealed the following (Supplement 8, Table S1):
- Completeness of data was present: Only 3 patients were lost to follow up and this did not cause risk of bias.
- Measurement instrument caused some concern about risk of bias: It was not clear if the instrument was used in a valid and reliable manner, but it was administered in the intended way using a valid representation of the outcome. It also appeared that the patients exhibited an understanding of the instrument that was used and did not encounter difficulties, but this was not reported.
- Data analysis did not cause concern for risk of bias: Adjustment, stratification, and model selection was appropriate based on an a-priority plan.

Overall risk of bias was deemed serious because of issue related to selection of participants into the study and the way the measurement instrument was used.

Figure 2. Rating individual signaling questions.

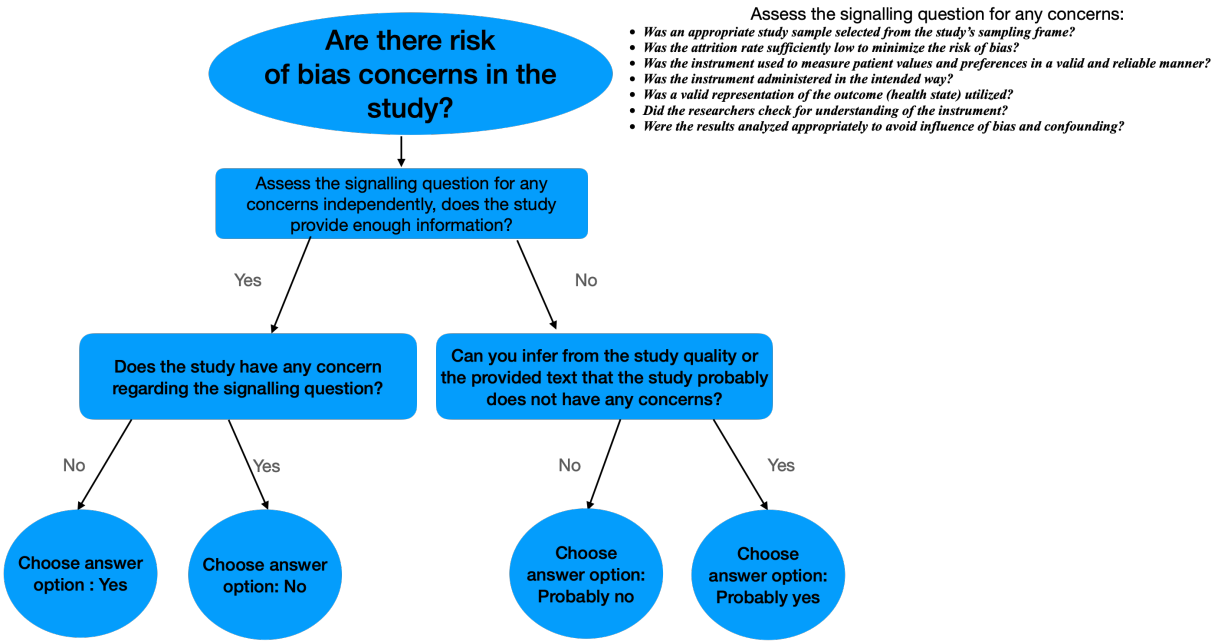


Table 4 Final judgment response

RESPONSE OPTION	CRITERIA
Not serious risk of bias;	The study is judged to have no serious risk of bias for all subdomains.
Serious risk of bias;	The study is judged to be serious risk of bias in at least one subdomain, but not very serious or extremely serious in any subdomain.
Very serious risk of bias (the study has some important problems);	The study is judged to be at very serious risk of bias in at least one subdomain, but not at extremely serious risk of bias in any subdomain.
Extremely serious risk of bias;	The study is judged to be at extremely serious risk of bias in at least one subdomain.

Discussion

We have developed and validated a new instrument to assess ROB in studies measuring the value, utility or relative importance that people place on health outcome, the ROBVALU tool. We followed a sequential mixed-methods approach, starting by adapting the signaling questions

from the GRADE guidance for judging the risk of bias across studies. ROBVALU differs from existing GRADE guidance in that it is specific for assessing ROB of individual studies as opposed to across studies [2]. We iteratively revised the tool with our core group and an advisory group. The final draft tool contains 15 items in four subdomains: selection of participants, completeness of data, measurement instrument, and data analysis. We conducted a validation exercise with 15 participants which showed good reliability. Additional refinement using a modified Delphi process established construct validity and the final content of the tool.

Assessing ROB is an essential step to assess the overall certainty of the evidence in a systematic review or health technology assessment and to develop a guideline. The assessment of ROB has often relied on adapting ROB tools not specifically designed for this type of research [27]. However, the lack of validation may lead to unreliable certainty of the evidence assessments, both for single studies and for a body of evidence. Using ROBVALUE, evaluators could incorporate the ROB assessment into their meta-analysis, such as performing a sensitivity analysis to evaluate how studies with higher risk of bias may affect the study's conclusion or primary outcomes. A particular advantage of the ROBVALU tool is that we used standardized GRADE terminology and judgments, that will facilitate assessing the ROB domain, when establishing the certainty of the evidence. Another advantage is that the ROBVALU tool can be used to assess ROB in all values utilities and importance of outcomes elicitation studies that utilize discrete choice, ranking, indifference, and rating methods [36]. It can also be used to assess ROB in individual studies that use indirect methods to elicit peoples preferences such as QoL and EQ-5D scores.

In addition to the strengths, this study, and the derived tool has also several limitations. The new tool focuses on assessing values quantitively. For any given intervention, there is usually qualitative literature exploring what patients want to achieve and what they value (or not) from interventions and this information may be important for decision-making. While some of the signaling questions may be used for qualitative studies, other signaling questions will not be applicable. Further exploration with qualitative studies should be performed to assess how ROBVALU may be adapted for that particular use case or whether a different tool is required. Another limitation of ROBVALU is the relatively poor fit of one of the items in our exploratory factor analysis, "*Was a valid representation of the outcome (health state) utilized?*", but this could be due to the relatively small sample size. However, we had a reason to retain this item based on the feedback from the Delphi panel who thought it was important. External validation of ROBVALU's reliability by different users and on different studies will help us refine the guidance, and to a smaller extent, the tool.

ROBVALUE allows appraising individual studies for their credibility and is not tied to using the GRADE approach. For example, in health technology assessments not using GRADE the certainty of input variables determines the certainty in decision-analytic model outputs, e.g., in cost utility and cost effectiveness analysis[33, 37]. ROBVALU should also be helpful when evaluating the ROB as part of a systematic review, health technology assessment, or a formal clinical health guideline, to develop recommendations and make judgments across the overall body of this type of evidence. That includes its use when following the GRADE approach, to assess the overall certainty of the evidence.

Ethics and Funding

This international study was designed and coordinated at McMaster University after approval by the Hamilton Integrated Research Ethics Board (HiREB) Project ID: 5634, and interviews and meetings were conducted in person or over video conference. All participants provided informed consent. The study was funded by the Canadian Institutes of Health Research (grant number 401310 to HJS).

Contributors and sources

Contributions of authors: SGK, YZ, JLB, and HJS conceived the project and were part of the core group. HJS oversaw the project. SGK, YZ, TD, JLB, HJS drafted the ROBVALU tool. JN, PAC, FX, and US were part of the advisory group. SGK led working groups and conducted the semi-structured interviews. SGK and LM analyzed the data. HPH, GH, YZ, and PAC were expert participants during study assessments. PAC, FX, BE, ZSP, VW, AS, JET, JN, LK, US were voting members in the Delphi process, and HJS, YZ, SGK, and JLB were non-voting members. SGK and HJS drafted the manuscript. YZ, JLB and HJS obtained funding for the study. All authors reviewed and commented on drafts of the manuscript.

Provenance

The authors are epidemiologists, statisticians, systematic reviewers, and health services researchers, many of whom are involved with in methods research and GRADE. Development of ROBVALU was informed by the GRADE guidelines 19, previously published tools for assessing risk of bias in intervention studies, systematic reviews of available tools to assess risk of bias in values and preferences, and by the authors' experience of developing similar tools to assess risk of bias. All authors contributed to development of ROBVALU tool and to writing associated guidance. All authors reviewed and commented on drafts of the manuscript. HJS will act as guarantor.

References

1. Boyd, C.M., et al., *Methods for benefit and harm assessment in systematic reviews.* 2012.
2. Zhang, Y., et al., *GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences—Risk of bias and indirectness.* Journal of clinical epidemiology, 2019. **111**: p. 94-104.
3. Pieterse, A.H. and A.M. Stiggelbout, *What are values, utilities, and preferences? A clarification in the context of decision making in health care, and an exploration of measurement issues.* Handbook of health decision science, 2016: p. 3-13.
4. Dolan, P., et al., *Valuing health states: a comparison of methods.* Journal of health economics, 1996. **15**(2): p. 209-231.
5. McDonough, C.M. and A.N. Tosteson, *Measuring preferences for cost-utility analysis: how choice of method may influence decision-making.* Pharmacoeconomics, 2007. **25**: p. 93-106.
6. Torrance, G.W., D. Feeny, and W. Furlong, *Visual analog scales: do they have a role in the measurement of preferences for health states?* Medical Decision Making, 2001. **21**(4): p. 329-334.
7. Rashidi, A.A., A.H. Anis, and C.A. Marra, *Do visual analogue scale (VAS) derived standard gamble (SG) utilities agree with Health Utilities Index utilities? A comparison of patient and community preferences for health status in rheumatoid arthritis patients.* Health and Quality of Life Outcomes, 2006. **4**: p. 1-10.
8. Bleichrodt, H. and M. Johannesson, *Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs.* Journal of health economics, 1997. **16**(2): p. 155-175.
9. Bridges, J.F., et al., *Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value in health, 2011. **14**(4): p. 403-413.
10. Horsman, J., et al., *The Health Utilities Index (HUI®): concepts, measurement properties and applications.* Health and quality of life outcomes, 2003. **1**(1): p. 1-13.
11. Devlin, N., D. Parkin, and B. Janssen, *Methods for analysing and reporting EQ-5D data*. 2020: Springer Nature.
12. Devlin, N., et al., *An introduction to EQ-5D instruments and their applications.* Methods for analysing and reporting EQ-5D data, 2020: p. 1-22.
13. Slaughter, K.B., et al., *Direct assessment of health utilities using the standard gamble among patients with primary intracerebral hemorrhage.* Circulation: Cardiovascular Quality and Outcomes, 2019. **12**(9): p. e005606.
14. Schünemann, H.J., et al., *The ecosystem of health decision making: from fragmentation to synergy.* The Lancet Public Health, 2022. **7**(4): p. e378-e390.
15. Zhang, Y., et al., *A systematic review of how patients value COPD outcomes.* Eur Respir J, 2018. **52**(1).
16. Etxeandia-Ikobaltzeta, I., et al., *Patient values and preferences regarding VTE disease: a systematic review to inform American Society of Hematology guidelines.* Blood advances, 2020. **4**(5): p. 953-968.
17. Alonso-Coello, P., et al., *GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction.* bmj, 2016. **353**.

18. Conrad, S., et al., *GRADE: Evidence to Decision (EtD) frameworks-a systematic and transparent approach to making well informed healthcare choices. 2: Clinical guidelines.* Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 2019. **140**: p. 63-73.
19. Sterne, J.A., et al., *RoB 2: a revised tool for assessing risk of bias in randomised trials.* bmj, 2019. **366**.
20. Sterne, J.A., et al., *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions.* bmj, 2016. **355**.
21. Morgan, R.L., et al., *A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE.* Environment International, 2019. **122**: p. 168-184.
22. Higgins, J., et al., *Risk Of Bias In Non-randomized Studies-of Exposure (ROBINS-E). Launch version, 1 June 2022*. 2022.
23. Wells, G.A., et al., *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* 2000.
24. Institute, J.B., *The Joanna Briggs Institute critical appraisal tools for use in JBI systematic reviews checklist for analytical cross sectional studies.* North Adelaide, Australia The Joanna Briggs Institute, 2017.
25. Mokkink, L.B., et al., *The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study.* Quality of life research, 2010. **19**: p. 539-549.
26. Sterne, J., *Higgins JPT RB on B of the DG for AN, Sterne J, Higgins J, Reeves B, on behalf of the development group for ACROBAT-NRSI.* A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI). Version, 2014. **100**.
27. Yepes-Nuñez, J.J., et al., *Forty-two systematic reviews generated 23 items for assessing the risk of bias in values and preferences' studies.* Journal of clinical epidemiology, 2017. **85**: p. 21-31.
28. Creswell, J.W. and V.L.P. Clark, *Designing and conducting mixed methods research*. 2017: Sage publications.
29. Helmer, O., *Analysis of the future: The Delphi method*. 1967, Rand Corp Santa Monica CA.
30. Nasa, P., R. Jain, and D. Juneja, *Delphi methodology in healthcare research: how to decide its appropriateness.* World Journal of Methodology, 2021. **11**(4): p. 116.
31. Murphy, M., et al., *Consensus development methods, and their use in clinical guideline development.* Health Technology Assessment (Winchester, England), 1998. **2**(3): p. i-88.
32. Siebert, U., et al., *State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force-3.* Value in Health, 2012. **15**(6): p. 812-820.
33. Siebert, U., *When should decision-analytic modeling be used in the economic evaluation of health care?* 2003, Springer. p. 143-150.
34. Kuehne, F., et al., *Causal analyses with target trial emulation for real-world evidence removed large self-inflicted biases: systematic bias assessment of ovarian cancer treatment effectiveness.* Journal of Clinical Epidemiology, 2022. **152**: p. 269-280.
35. Goossens, L.M., et al., *Is the EQ-5D responsive to recovery from a moderate COPD exacerbation?* Respiratory medicine, 2011. **105**(8): p. 1195-1202.

36. Soekhai, V., et al., *Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review.* Drug discovery today, 2019. **24**(7): p. 1324-1331.
37. Caro, J.J., et al., *Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1.* Medical Decision Making, 2012. **32**(5): p. 667-677.
38. Bøgelund, M., et al., *Patient preferences for diabetes management among people with type 2 diabetes in Denmark–a discrete choice experiment.* Current medical research and opinion, 2011. **27**(11): p. 2175-2183.
39. Brown, S.E., et al., *Perceptions of Quality of Life Effects of Diabetes Treatments Among Vulnerable and Non-Vulnerable Older Patients Running title: Perceptions of Diabetes Treatments.* Journal of the American Geriatrics Society, 2008. **56**(7): p. 1183.
40. Jendle, J., et al., *Willingness to pay for diabetes drug therapy in type 2 diabetes patients: based on LEAD clinical programme results.* Journal of Medical Economics, 2012. **15**(sup2): p. 1-5.
41. Watson, V., et al., *Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia.* The Journal of urology, 2004. **172**(6): p. 2321-2325.
42. Llewellyn-Thomas, H.A., et al., *Using a trade-off technique to assess patients' treatment preferences for benign prostatic hyperplasia.* Medical Decision Making, 1996. **16**(3): p. 262-272.
43. Piercy, G.B., et al., *Impact of a shared decision-making program on patients with benign prostatic hyperplasia.* Urology, 1999. **53**(5): p. 913-920.
44. Bossema, E., et al., *Patients' preferences for low rectal cancer surgery.* European Journal of Surgical Oncology (EJSO), 2008. **34**(1): p. 42-48.
45. Harrison, J.D., et al., *Patient and physician preferences for surgical and adjuvant treatment options for rectal cancer.* Archives of Surgery, 2008. **143**(4): p. 389-394.
46. Solomon, M.J., et al., *What do patients want? Patient preferences and surrogate decision making in the treatment of colorectal cancer.* Diseases of the Colon & Rectum, 2003. **46**: p. 1351-1357.
47. Blinman, P., et al., *Adjuvant chemotherapy for early colon cancer: what survival benefits make it worthwhile?* European journal of cancer, 2010. **46**(10): p. 1800-1807.
48. Zolciak, A., et al., *Abdominoperineal resection or anterior resection for rectal cancer: patient preferences before and after treatment.* Colorectal Disease, 2006. **8**(7): p. 575-580.
49. Lenert, L.A. and R.M. Soetikno, *Automated computer interviews to elicit utilities: potential applications in the treatment of deep venous thrombosis.* J Am Med Inform Assoc, 1997. **4**(1): p. 49-56.
50. Thomson, R., et al., *Decision analysis and guidelines for anticoagulant therapy to prevent stroke in patients with atrial fibrillation.* Lancet, 2000. **355**(9208): p. 956-62.
51. Polonsky, W.H., et al., *Patient perspectives on once-weekly medications for diabetes.* Diabetes Obes Metab, 2011. **13**(2): p. 144-9.
52. Arnesen, T. and M. Trommald, *Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method.* J Health Serv Res Policy, 2004. **9**(1): p. 43-50.
53. Arnesen, T. and M. Trommald, *Are QALYs based on time trade-off comparable?--A systematic review of TTO methodologies.* Health Econ, 2005. **14**(1): p. 39-53.

54. Craig, B.M., J.J. Busschbach, and J.A. Salomon, *Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods.* Med Care, 2009. **47**(6): p. 634-41.
55. Lin, M.R., et al., *Rating scale, standard gamble, and time trade-off for people with traumatic spinal cord injuries.* Phys Ther, 2006. **86**(3): p. 337-44.
56. Gage, B.F., et al., *Cost-effectiveness of warfarin and aspirin for prophylaxis of stroke in patients with nonvalvular atrial fibrillation.* JAMA, 1995. **274**(23): p. 1839-45.
57. Holmberg, M.J. and L.W. Andersen, *Collider bias.* Jama, 2022. **327**(13): p. 1282-1283.
58. Hol, L., et al., *Preferences for colorectal cancer screening strategies: a discrete choice experiment.* Br J Cancer, 2010. **102**(6): p. 972-80.