The European Journal of Health Economics
https://doi.org/10.1007/s10198-021-01286-0

ORIGINAL PAPER

Evaluating the conduct and application of health utility studies: a review of critical appraisal tools and reporting checklists

Michael J. Zoratti[1] · A. Simon Pickard[2] · Peep F. M. Stalmeier[3] · Daniel Ollendorf[4] · Andrew Lloyd[5] · Kelvin K W Chan[6,7] · Don Husereau[8] · John E. Brazier[9] · Murray Krahn[10] · Mitchell Levine[1,11] · Lehana Thabane[1,12] · Feng Xie[1] on behalf of Health Utility Book (HUB) Working Group

Received: 22 July 2020 / Accepted: 12 March 2021

Abstract

Background Published health utility studies are increasingly cited in cost–utility analyses to inform reimbursement decision-making. However, there is limited guidance for investigators looking to systematically evaluate the methodological quality of health utility studies or their applicability to decision contexts.

Objective To describe how health utility concepts are reflected in tools intended for use with the health economic literature, particularly with respect to the evaluation of methodological quality and context applicability.

Methods We reviewed the critical appraisal and reporting tools described in a 2012 report published by the Agency for Healthcare Research and Quality (AHRQ), supplemented with a keyword search of MEDLINE and EMBASE, to identify existing tools which include health utility constructs. From these tools, a list of relevant items was compiled and grouped into domain categories based on the methodological or applicability aspect they were directed toward.

Results Of the 24 tools we identified, 12 contained items relevant to the evaluation of health utilities. Sixty-five items were considered relevant to the evaluation of quality, while 44 were relevant to the evaluation of applicability. Items were arranged into four domains: health state descriptions; selection and description of respondents; elicitation and measurement methods; and other considerations.

Conclusion As key inputs to cost–utility analyses, health utilities have the potential to significantly impact estimates of cost-effectiveness. Existing tools contain only general items related to the conduct or use of health utility studies. There is a need to develop tools that systematically evaluate the methodological quality and applicability of health utility studies.

Keywords Health utility · Checklist · Cost utility analysis · Methodology

Introduction

Cost–utility analyses (CUAs) are referenced in the reimbursement decision-making processes of many jurisdictions globally. One of the major strengths of the CUA is its generic outcome measure, the quality-adjusted life year (QALY), which facilitates comparisons across diseases and interventions. Though they did not explicitly use the term QALY, Klarman and colleagues first described the concept of reflecting both quantity and quality of life in a single index [1]. In CUAs, health-related quality of life is quantified with health utilities which reflect cardinal preferences for health states, anchored at 0 for dead and 1 for full health. Higher health state utility values (HSUVs) reflect better health status, while negative values indicate health states worse than dead. Multi-attribute utility instruments (MAUIs) have become a widely used approach to measure HSUVs.

Despite the advantages of using HSUVs to value health outcomes, investigators must be aware of the significant variation in estimates for health utilities across studies and populations [2–11]. A study by Richardson and colleagues, who compared six MAUIs (EQ-5D-5L, SF-6D, HUI 3, 15D, QWB, and AQoL-8D), concluded that these instruments measure similar but different constructs, with variations attributed to differences in the questions or response categories used to describe health states [11]. These instruments also differ with respect to the range of plausible values. For example, the EQ-5D-3L generates estimates from

✉ Feng Xie
fengxie@mcmaster.ca

Extended author information available on the last page of the article

Published online: 11 April 2021

-0.59 to + 1.00 using the United Kingdom value set [12], while the SF-6Dv2 values vary from -0.574 to + 1.00 [13].

There are good reasons for the differences that are observed between the values obtained from the various MAUIs. Some key sources of differences include the way health is characterized based on the description system, the valuation approaches to elicit preferences, the mode of elicitation of preferences (e.g., online vs face-to-face), and respondent characteristics [14–16]. However, some studies may not use appropriate methods for valuation. For instance, in a systematic review of health utilities in lung cancer, Sturza and colleagues reported that 16% of estimates (35/223, 7 studies) were derived using simple judgment, rather than established, valid preference elicitation techniques [3]. Irrespective of the source of variation, these differences introduce additional uncertainty to cost-effectiveness models, and, in some cases, these differences have the potential to influence reimbursement decisions.

Health utilities are playing an increasingly important role in informing reimbursement decision-making and, consequently, are increasingly receiving attention from the health economics and outcomes research community. In 2014, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) established a task force for the study of good practices for investigators using health utilities in economic evaluations. In a recent publication, the task force acknowledged that the selection of published health utility inputs for CUAs is commonplace in the absence of primary data, though the means by which investigators select and appraise the literature are often not ideal [17]. However, the so-called ideal evidence might not be available in the literature despite the conduct of robust and comprehensive searches. In such cases, investigators must weigh considerations of data quality against data appropriateness.

To support minimum reporting standards in the selection of valid health utility inputs for cost-effectiveness models, the task force proposed the ISPOR HSU Good Practice Task Force Minimum Reporting Standards of Systematic Review of Utilities for Cost-Effectiveness models (ISPOR SpRUCE checklist). As part of the review process in selecting health utility inputs, the checklist presents an item each for a quality check ("Describe the quality criteria used during the review to decide whether to include or exclude studies from the analysis") and the assessment of health state utility relevance to the decision context ("Describe the relevance of HSUs to the cost-effectiveness model and the target reimbursement agency if appropriate"). While the concepts are elaborated on in the text, the checklist itself stops short of listing any specific criteria. Thus, there is an acknowledged need for a systematic and comprehensive approach to the identification, appraisal, and selection of health utilities.

Importantly, there are a lack of tools available to investigators to support the direct and explicit evaluation of health utility studies, both with respect to their quality and their relevance or applicability to a given decision problem [18]. Here, quality refers to methodological robustness, or whether a study was conducted in accordance with good practices, including how respondents are sampled, how evaluations are administered, and how data are analyzed. Applicability is meant to refer to the relevance of a given health utility estimate to a specific decision context. This may include considerations of differences in clinical characteristics, differences in the demographic makeup of the study sample compared to the patients whose outcomes are being modeled in an economic evaluation, and whether the health utility study's methods are aligned with the requirements or guidelines of a given decision-making body. To support the development of two novel tools for the assessment of health utility study quality and applicability, we present a review of existing quality assessment and reporting tools or checklists in the health economics and health utility literature.

Methods

In 2012, the United States-based Agency for Healthcare Research and Quality (AHRQ) conducted a systematic literature review to evaluate best practices for conducting and reporting health economic evaluations. This review identified ten quality assessment tools for economic evaluations published between 1992 and 2011 [19]. To identify additional instruments, such as those published after 2012, we supplemented this review with keyword search of medical literature databases, including MEDLINE and EMBASE (via Ovid). Keyword searches included "health utility" combined with "quality", "appraisal", "applicability", or "relevance". Hand searches of the reference lists of recently published systematic reviews of economic literature were also conducted to identify means by which authors evaluated the quality of the included studies. Searches were conducted to March 1, 2020.

Two reviewers (MZ and FX), working independently and in duplicate, reviewed all items in each instrument for relevance in evaluating the methodological quality of a health utility study or the applicability of the health utility study to the context of the cost-effectiveness model. Differences in item selection were resolved by discussion.

Tools intended for use with economic evaluations or with health utility studies were eligible for inclusion if they contained items relevant to health utility measures, including evaluating health utility study quality or decision-context applicability. Instruments must have been published in English. Items were considered related to quality if they were directed to the methods used in the study or the way in which the study was conducted, while items were considered related to applicability if they were directed toward

evaluating the context or relevance of the health utility estimates to the decision problem studied in an economic evaluation.

A list of relevant items was compiled and grouped into categories based on the methodological or applicability aspect they were relevant to evaluating (i.e., the selection of respondents, health states, preference elicitation and measurement methodologies, and other elements). Items could be labeled under both categories. Additionally, we collected information on the primary purpose of the tools (i.e., reporting checklist and/or critical appraisal), the response options, and the methods by which the tool was developed.

Results

Summary of the literature review

We identified and reviewed the content of 24 tools, checklists, and frameworks [17, 20–44]. Of these, 12 [17, 20, 22, 23, 25–30, 41–44] contained items which are considered potentially relevant to evaluating the methodological quality and applicability of health utility studies (Fig. 1, Table 1). The 12 tools were diverse, varying with respect to primary purpose, response options, methods of development, and number of items. One tool prompts users to come to an overall study score based on a weighted scoring system, while others presented binary yes–no response options, multiple response options, or free text. Most tools were developed through expert panels or consultations. The number of items varied from a minimum of 15 to a maximum of 57. Together, the 12 tools consisted of 354 items, 65 of which are considered potentially relevant to the evaluation of study quality and 44 items relevant to applicability (Table 2). These items were arranged into four general domains: health state descriptions; selection and description of respondents; elicitation and measurement methods; and other considerations.

Tools for health economic evaluation appraisal and reporting ($n = 6$)

Out of the 12 tools included, six were designed for use at the economic evaluation level, rather than with a specific focus on health utility (Tables A1–A6) [20, 23, 29, 30, 41–44]. Collectively, there are 10 items potentially relevant for assessing quality and 10 for applicability. These items were high-level, non-specific references to the source of health utilities used in the economic evaluation, means of obtaining, description, or value weights of health utilities. Overall, these items are considered potentially relevant only at a strictly conceptual level.

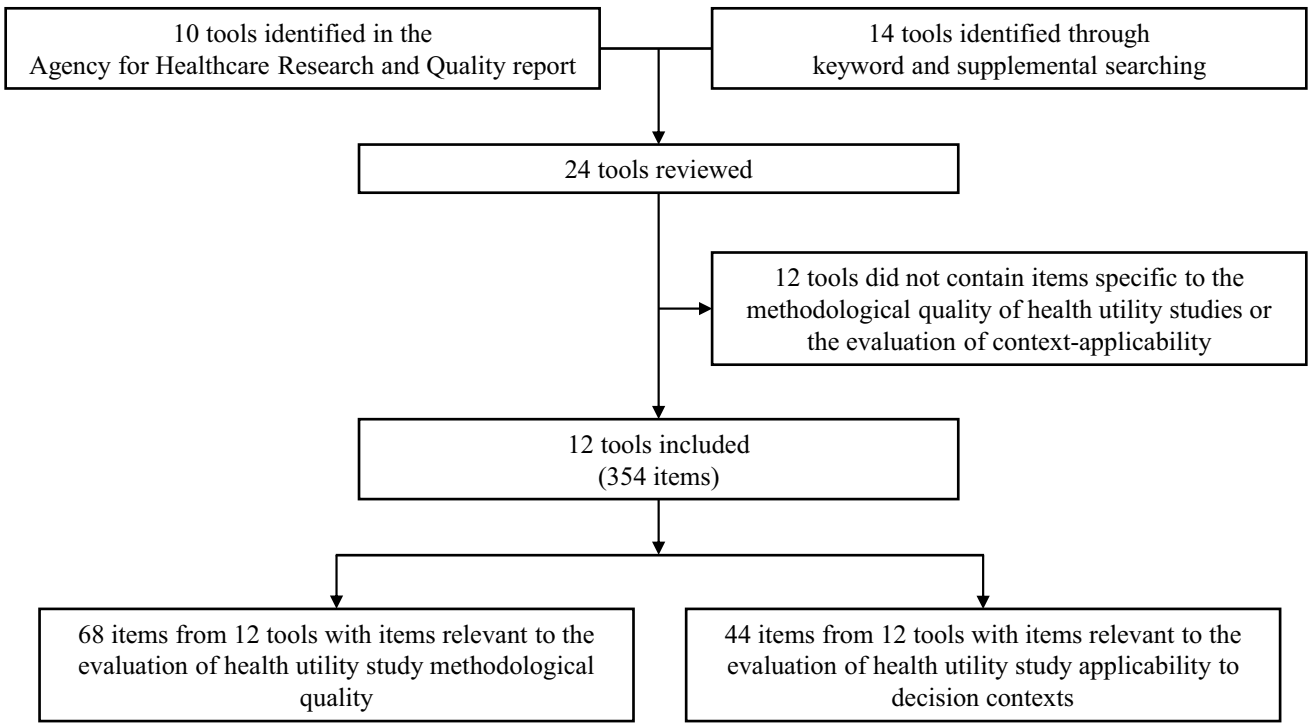

Fig. 1 Process of selecting tools and relevant item

Table 1 Summary of 12 tools included in the review

Author/year	Tool name	Primary purpose	Item response options	Development method	Number of items	Number of relevant items	
						Quality ($n=65$)	Applicability ($n=44$)
Tools for health economic evaluations							
Drummond [29]	British Medical Journal (BMJ) Checklist	Reporting checklist	Yes/No	Expert consultation + survey to achieve a broad consensus	35	2	2
Chiou [23]	Quality of Health Economic Studies (QHES) Instrument	Reporting checklist + Critical appraisal	Weighted scoring	Expert panel	16	1	1
Ungar [30]	Pediatric Quality Appraisal Questionnaire (PQAQ)	Reporting checklist + Critical appraisal	Multiple response options	Review of existing tools + Expert panel	57	3	3
First US CEA Panel [41–43]	–	Reporting checklist	Yes/No	Expert panel	37	1	2
Husereau [20]	Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement	Reporting checklist	Indicate page numbers[a]	Expert panel	24	1	1
Second US CEA Panel [44]	–	Reporting checklist	Yes/no	Expert panel	45	2	1
Tools specifically for health utility studies							
Brazier [28]	–	Critical appraisal	Open response[b]	Author-proposed	24	3	5
Stalmeier [25]	–	Reporting checklist	Yes/no	Author-proposed + Expert panel	40	15	2
Petrou [26]	MApping onto Preference-based measures reporting Standards (MAPS) statement	Reporting checklist	Indicate page numbers[a]	Expert panel	23	15	2
Xie [27]	Checklist for REporting VAluaTion StudiEs (CREATE)	Reporting checklist	Yes/no	Expert panel	21	13	3
Nerich [22]	–	Reporting checklist + critical appraisal	Yes/no Open response[b]	Author-proposed	17	4	12
Brazier [17]	ISPOR Minimum Reporting Standards of Systematic Review of Utilities for Cost-Effectiveness models (ISPOR SpRUCE Checklist)	Reporting checklist	Open response[b]	Expert panel	15	5	10

Reporting checklists are used to evaluate the presence or absence of components without value on that component's use. Critical appraisal tools are an extension of reporting checklists and include some interpretation or evaluation of the reported content

[a]Instruments where response options include a requirement to list the page number(s) corresponding to the item's criteria

[b]Instruments with no pre-specified response options and where investigators instead provide a free-text response to the instrument's items

Table 2 Items of relevance for assessing quality and applicability of health utility studies

Item	Quality ($n=65$*)	Applicability ($n=44$*)
Health state descriptions		
Complete description of estimates of effectiveness, resource use, unit costs, health states, and quality of life weights and their sources [41–43]	✓	✓
Description of health states, if any [25]	✓	✓
How is "perfect health" described? [25]	✓	
How is "worst health" described? [25]	✓	
If utility for "own health" was assessed, was own health specified further? [25]	✓	
Description of treatments, if any (a treatment corresponds to a decision option in a decision tree) [25]	✓	
Selection and description of respondents		
Whose quality of life is assessed? [30]	✓	✓
Details of the subjects from whom evaluations were obtained are given [29]	✓	✓
Is a comprehensive description provided for the population used to elicit HSUVs (i.e., characteristics, size, and nationality)? [22]	✓	✓
Is an explanation provided for the choice of the population used to elicit HSUVs (i.e., patient, healthcare professional [and type], expert, general population)? [22]	✓	✓
Response rate for the measure used [17]	✓	
Extent of missing data or data lost to follow-up [17]	✓	
Are the HSUVs appropriate with respect to comparability of populations (i.e., diagnosis and disease severity)? [22]		✓
Are the HSUVs appropriate with respect to comparability of countries? [22]		✓
Population or patient characteristics [17]		✓
Are the items relevant and appropriate for the population? [28]		✓
Whose values have been used? [28]		✓
How well are the preferences of the patients/general population/decision-makers likely to conform to these assumptions? [28]		✓
Are the background characteristics of the respondents to the valuation survey representative of the population? [28]		✓
Measurement and valuation of preference-based outcomes (If applicable, describe the population and methods used to elicit preferences for outcomes) [20]	✓	✓
Describe how the estimation sample was identified, why it was selected, the methods of recruitment and data collection, and its location(s) or setting(s) [26]	✓	✓
If an external validation sample was used, the rationale for selection, the methods of recruitment and data collection, and its location(s) or setting(s) should be described [26]	✓	✓
State the size of the estimation sample and any validation sample(s) used in the analyses (including both number of individuals and number of observations) [26]	✓	
Describe the characteristics of individuals in the sample(s) (or refer back to previous publications giving such information). Provide summary scores for source and target measures, and summarize results of analyses used to assess overlap between the source and target measures [26]	✓	
Sample size/power calculations are stated and rationalized [27]	✓	✓ ✓
Target population is described [27]	✓	
Sampling method is stated and rationalized [27]	✓	
Recruitment strategies are described [27]	✓	
Response rate is reported [27]	✓	
Reasons for excluding any respondents or observations are provided [27]	✓	
Characteristics of respondents included in the analysis are described [27]	✓	✓
Elicitation and measurement methods		
Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for the measures/scales used? [23]	✓	✓
Is an explanation provided for the choice of technique(s) used to elicit HSUVs? [22]	✓	✓
If quality of life is measured, what type of instrument is used? [30]	✓	✓
Methods to value health states and other benefits are stated [29]	✓	✓

Table 2 (continued)

Item	Quality ($n=65$*)	Applicability ($n=44$*)
Is a comprehensive description provided of technique(s) used to elicit the obtained HSUVs? [22]	✓	✓
If more than one utility measure was used, was the presentation order randomized? If not, what was the order? [25]	✓	
Who performed the quality-of-life assessment? [30]	✓	✓
How were the utility questions administered (e.g., by interview, mailed questionnaires, computer, the Internet, or self-administered under general supervision)? If by interview, how were interviewers trained? [25]	✓	
Was the presentation order of the health states randomized? If not, what was the order? [25]	✓	
Was it made explicit that each duration was followed by death? [25]	✓	
Was the subject instructed to assume that survival does not occur with knowledge of the date of death? [25]	✓	
Was the health state labeled or unlabeled? [25]	✓	
Were subjects confronted with inconsistencies in their scores, such as a change in the health state ordering as inferred from the different utility assessment methods? [25]	✓	
Was a matching or choice indifference search procedure used? If yes, which particular indifference search procedure was used? [25]	✓	
Which software program (if any) was used? Was it used by the subject alone, or was someone present in the start-up phase to answer questions or detect misconceptions? If someone was present, how was he or she trained? [25]	✓	
What visual aids, if any, were used (e.g., rulers, pies, probability wheels, or other means of visualizing probabilities or trade-offs)? [25]	✓	
What is the model of preferences being assumed? [28]	✓	
What are the main assumptions of this model? [28]	✓	
Which choice-based method has been used? [28]	✓	
Measure used [17]	✓	✓
Preference weights [17]	✓	✓
Methods for obtaining estimates of effectiveness, costs and preferences [41–43]		✓
Which method was chosen (e.g., visual analogue scale, time trade-off, standard gamble, willingness to pay)? [25]		✓
Does the instrument cover all dimensions of health of interest? [28]		✓
Descriptive statistics about HSUs [17]		✓
Measurement and valuation of preference-based outcomes (If applicable, describe the population and methods used to elicit preferences for outcomes) [20]	✓	✓
Describe how predicted scores or utilities are estimated for each model specification [26]	✓	
Methods for obtaining estimates of costs and preference weights [44]	✓	✓
The attributes of the system are described [27]	✓	
The number of levels in each attribute of the instrument is described [27]	✓	
Method(s) of assigning the health states to respondents are stated [27]	✓	
Mode of data collection is stated [27]	✓	✓
Preference elicitation technique(s) are described [27]	✓	
Other considerations		
Is the difference between when the CUA was performed and when the HSUVs were elicited less than 10 years? [22]		✓
Do the authors use the same HSUVs in the CUA as presented in the original data source? [22]		✓
Do the authors discuss the limitations of the data source selection, the elicitation, and the use of HSUVs in the CUA? [22]		✓
Original reference [17]		✓
Actual HSUs used [17]		✓
Adjustments or assumptions [17]		✓
Describe the source and target measures and the methods by which they were applied in the mapping study [26]	✓	
Describe the methods used to assess the degree of conceptual overlap between the source and target measures [26]	✓	
State how much data were missing and how missing data were handled in the sample(s) used for the analyses [26]	✓	
Describe and justify the statistical model(s) used to develop the mapping algorithm [26]	✓	
Describe and justify the methods used to validate the mapping algorithm [26]	✓	
State and justify the measure(s) of model performance that determine the choice of the preferred model(s) and describe how these measures were estimated and applied [26]	✓	

Table 2 (continued)

Item	Quality ($n=65$*)	Applicability ($n=44$*)
State which model(s) is(are) preferred and justify why this(these) model(s) was(were) chosen [26]	✓	
Provide all model coefficients and standard errors for the selected model(s). Provide clear guidance on how a user can calculate utility scores based on the outputs of the selected model(s) [26]	✓	
Report information that enables users to estimate standard errors around mean utility predictions and individual-level variability [26]	✓	
Present results of model performance, such as measures of prediction accuracy and fit statistics for the selected model(s) in a table or in the text. Provide an assessment of face validity of the selected model(s) [26]	✓	
Critique of data quality [44]	✓	
The approach to selecting health states to be valued directly is explained [27]	✓	

The SpRUCE Checklist includes items directing investigators to perform a quality check of the health utility study, to assess the relevance of the health state utilities to the cost-effectiveness model and the target reimbursement agency, and to present the rationale for selecting the health states utilities used in the cost-effectiveness model (3 items). Similarly, the checklist by Nerich and colleagues asks investigators to identify the data sources of health utility estimates and to provide a description and explanation for explicit assumptions made in the use of health utility values in the CUA (3 items)

*Some multi-component questions have been collapsed into a single item

Tools specifically for health utility measures ($n=6$)

The remaining six tools (Tables A7–A12), together comprising 140 items, were developed specifically toward the evaluation of health utility studies [17, 22, 25–28]. From these instruments, 54 items are considered relevant to the evaluation of study quality and 33 for applicability. A brief description of each of the six tools is provided below.

Brazier et al. (1999)

In a 1999 review of the use of health measures in economic evaluations, Brazier and colleagues presented a checklist for judging the merits of preference-based measures of health (Table A7) [28]. This 24-item, author-proposed checklist is arranged into five major categories to address practicality, reliability, and three aspects of validity. Across categories, three and five items, respectively, are considered potentially relevant to assess the quality and applicability of published health utility studies.

Stalmeier et al. (2001)

Stalmeier and colleagues published a reporting checklist of essential items to guide the drafting of methods sections of health utility studies (Table A8) [25]. This 40-item tool was developed through consultation with a panel of eight experts. The tool comprises six main reporting categories: design; administration; health state descriptions; description of the utility assessment method; indifference procedures; and visual aids and software programs. Fifteen items are potentially relevant to evaluating quality and two items to applicability ("Description of health states, if any"; "Which method was chosen [e.g., visual analogue scale, time tradeoff, standard gamble, willingness to pay]?").

MAPS (2015)

Mapping methods, where investigators develop and apply an algorithm to non-utility data to predict health utility values, have been used to indirectly estimate health utilities in recent years. Petrou and colleagues developed the 23-item MApping onto Preference-based measures reporting Standards (MAPS) statement (Table A9) [26]. This reporting checklist is intended to promote complete and transparent reporting and its items are arranged to reflect the components of a traditional journal manuscript. All MAPS statement items pertaining to reporting of methods and results are considered potentially relevant to the evaluation of study quality. Two methods items are also considered potentially relevant to the evaluation of applicability ("Describe how the estimation sample was identified, why it was selected, the methods of recruitment and data collection, and its location(s) or setting(s)"; "If an external validation sample was used, the rationale for selection, the methods of recruitment and data collection, and its location(s) or setting(s) should be described").

CREATE 2015

The 21-item Checklist for REporting VAluaTion StudiEs (CREATE) instrument was developed through a modified, two-round Delphi panel to promote good reporting practices and serve as a guide for investigators engaged in valuation studies (Table A10) [27]. Twelve items pertaining to the descriptive system, health states valued, sampling,

preference data collection, and study sample are considered potentially relevant to the evaluation of study quality. Three items are relevant to the evaluation of applicability ("Target population is described"; "Mode of data collection is stated"; "Characteristics of respondents included in the analysis are described").

Nerich et al. 2017

Our review found one tool specific to the critical appraisal of the health utility literature (Table A11) [22]. Citing a lack of a means to critically appraise published health utility studies identified in a systematic literature review, Nerich and colleagues proposed a 3-part, 17-item checklist to evaluate HSUVs applied to breast cancer CUAs in terms of the data source (3 items), the elicitation method (4 items), and the application of the study in economic evaluations (10 items). We considered 4 items on the elicitation of HSUVs to be potentially relevant to the quality assessment, with an additional 12 items potentially relevant to applicability.

SpRUCE Checklist 2019

The concepts presented in the recently published SpRUCE checklist are directly relevant to our research question, with specific items directing investigators to assess and report on study quality and relevance (Table A12) [17]. This 15-item checklist lists several concepts that should be reported to support the selection of health utility inputs. As the checklist is intended to guide reporting, however, there are opportunities to develop specific criteria to shape practice. For instance, a single item highlights the need to report an assessment of relevance ("Assessment of HSU relevance"), though the checklist does not provide further elaboration on the concepts to consider.

Discussion

As influential inputs in CUAs, published health utilities have the potential to significantly impact estimates of cost effectiveness [4, 10, 11]. In the context of reimbursement decision-making, this may sway pricing and reimbursement policy and funding allocations. However, little research has been conducted into means by which we may assess the methodological robustness of health utility studies or evaluate their applicability to reimbursement decision-making contexts.

The tools identified in this review were comprehensive for their intended purposes and provide investigators with a means of thoroughly evaluating the health economic literature. The majority were developed through consultation with expert panels, which lends credibility and weight to the importance of the constructs they include. However, most tools were directed to the conduct of health economic evaluations and, given the scope of evidence to be weighed for these study designs, it is reasonable for many parameters to only be considered at a high level. Relatively few items from any tool were considered fit for evaluating either methodological quality or applicability considerations, with many tools only including a single relevant item for either category. There are some important exceptions to this observation, where tools incorporated several items directed to various aspects of either quality or applicability [17, 22, 25–27]. Yet, these are not without limitations.

Collectively, these tools share a common objective of improving the transparency and clarity in the reporting of economic evaluations or health utility studies. However, most have been conceptualized as reporting checklists. While this format encourages some degree of scrutiny and comparison across the literature, they are limited in their utility for explicitly assessing either methodological rigor or applicability considerations. A focus on reporting rather than appraisal may not fully support an investigator's need to select the best available evidence. Rather, tools designed to directly engage investigators and reviewers in appraising study methods or applicability encourage further, critical engagement with the literature. By considering key elements that may differentiate a well-conducted study from one with important methodological limitations, for instance, investigators may be better positioned to make parameter selections that are robust, valid, and defendable. Such tools are commonplace for the clinical literature, such as the Cochrane Collaboration's tool for assessing risk of bias in randomized trials [45], yet are critically lacking for evaluations of the health utility literature.

We observed considerable overlap in the constructs represented across the 12 tools, with most items directed to elicitation methods. Several items are related to health state descriptions, respondent characteristics, and elicitation and measurement methods. These concepts are broadly relevant to the purposes of evaluating the methodological quality of a health utility study and, taken together, provide a strong framework upon which to propose a novel instrument. Most of the items we considered potentially relevant to assessing applicability are directed at evaluating the selection and descriptions of respondents or elicitation and measurement methods. When evaluating the relevance of these items, it is important to consider that most of the tools we identified were not developed specifically for the purposes of assessing health utility studies. Indeed, the included items vary with respect to their relevance and significance for the objectives of the proposed quality and applicability tools. Similarly, items also vary in terms of granularity, from high-level questions regarding the names of measures used to collect health utility data to the provision of verbatim health state

descriptions. We also observe overlap in items flagged for quality and relevance. This reflects the shared or complementary purposes of the tools considered.

Beyond the reporting checklists and appraisal tools, which focus primarily on explicitly described study content, there are other considerations that deserve weight. For example, the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) checklist, developed through a Delphi study with 57 participating experts, was proposed as a means of evaluating the methodological quality of studies reporting on the measurement properties of patient-reported outcomes [46]. This focus on psychometric properties is largely absent in commonly used checklists for the health economic literature yet addresses the validity of the instruments used to generate the health utility estimates upon which CUAs rely. As this checklist is not specific to the evaluation of health utilities, it was not considered eligible for formal inclusion in this review.

Conclusion

Existing checklists or appraisal tools for health economic evaluations contain some general items related to the conduct or use of health utility studies. However, there lacks a tool to guide the systematic evaluation of the quality and applicability of published health utilities in the context of coverage or reimbursement decision-making. Thus, there lies an opportunity to expand on the current methods literature with novel tools to complement existing guidance. Through this review, we have described existing frameworks intended for use with the health economics literature and have identified data elements specifically relevant to the evaluation of credibility and applicability. These efforts are not directed toward determining the relative value or integrity of one health utility measure over another, but rather are focused on establishing some criteria against which investigators may critically appraise a study's methodology and arrive at a conclusion concerning its robustness and validity. This work will directly support the development of two new tools to promote transparency, accountability, and methodological rigor in the application of the health utility literature in decision-making.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10198-021-01286-0.

Funding No funding was received for this research.

Declarations

Conflicts of interest None to declare.

References

1. Klarman, H.E., Francis, J.O.S., Rosenthal, G.D.: Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. Med. Care **6**(1), 48–54 (1968)
2. Bremner, K.E., Chong, C.A., Tomlinson, G., Alibhai, S.M., Krahn, M.D.: A review and meta-analysis of prostate cancer utilities. Med. Decis. Making **27**(3), 288–298 (2007). https://doi.org/10.1177/0272989x07300604
3. Sturza, J.: A review and meta-analysis of utility values for lung cancer. Med. Decis. Making **30**(6), 685–693 (2010). https://doi.org/10.1177/0272989x10369004
4. Paracha, N., Thuresson, P.O., Moreno, S.G., MacGilchrist, K.S.: Health state utility values in locally advanced and metastatic breast cancer by treatment line: a systematic review. Expert Rev. Pharmacoecon. Outcomes Res. **16**(5), 549–559 (2016). https://doi.org/10.1080/14737167.2016.1222907
5. Peasgood, T., Ward, S.E., Brazier, J.: Health-state utility values in breast cancer. Expert Rev. Pharmacoecon. Outcomes Res. **10**(5), 553–566 (2010). https://doi.org/10.1586/erp.10.65
6. Hao, Y., Wolfram, V., Cook, J.: A structured review of health utility measures and elicitation in advanced/metastatic breast cancer. Clin. Econ. Outcomes Res. CEOR **8**, 293–303 (2016). https://doi.org/10.2147/ceor.s100448
7. Schiller-Fruhwirth, I.C., Jahn, B., Arvandi, M., Siebert, U.: Cost-effectiveness models in breast cancer screening in the general population: a systematic review. Appl. Health Econ. Health Policy **15**(3), 333–351 (2017). https://doi.org/10.1007/s40258-017-0312-3
8. Carter, G.C., King, D.T., Hess, L.M., Mitchell, S.A., Taipale, K.L., Kiiskinen, U., Rajan, N., Novick, D., Liepa, A.M.: Health state utility values associated with advanced gastric, oesophageal, or gastro-oesophageal junction adenocarcinoma: a systematic review. J. Med. Econ. **18**(11), 954–966 (2015). https://doi.org/10.3111/13696998.2015.1066380
9. Djalalov, S., Rabeneck, L., Tomlinson, G., Bremner, K.E., Hilsden, R., Hoch, J.S.: A review and meta-analysis of colorectal cancer utilities. Med. Decis. Making **34**(6), 809–818 (2014). https://doi.org/10.1177/0272989x14536779
10. Jeong, K., Cairns, J.: Systematic review of health state utility values for economic evaluation of colorectal cancer. Heal. Econ. Rev. **6**(1), 36 (2016). https://doi.org/10.1186/s13561-016-0115-5
11. Richardson, J., Khan, M.A., Iezzi, A., Maxwell, A.: Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. Med. Decis. Making **35**(3), 276–291 (2015). https://doi.org/10.1177/0272989x14543107
12. Dolan, P.: Modeling valuations for EuroQol health states. Med. Care **35**(11), 1095–1108 (1997). https://doi.org/10.1097/00005650-199711000-00002
13. Mulhern, B.J., Bansback, N., Norman, R., Brazier, J.: Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. Med. Care **58**(6), 566–573 (2020). https://doi.org/10.1097/mlr.0000000000001324
14. Galante, J., Augustovski, F., Colantonio, L., Bardach, A., Caporale, J., Marti, S.G., Kind, P.: Estimation and comparison of EQ-5D health states' utility weights for pneumoccocal and human papillomavirus diseases in Argentina, Chile, and the United Kingdom. Value Health **14**(5, Supplement), S60–S64 (2011). https://doi.org/10.1016/j.jval.2011.05.007
15. Takemoto, M.L., Lopes da Silva, N., Ribeiro-Pereira, A.C., Schilithz, A.O., Suzuki, C.: Differences in utility scores obtained

through Brazilian and UK value sets: a cross-sectional study. Health Qual. Life Outcomes **13**, 119 (2015). https://doi.org/10.1186/s12955-015-0318-1
16. Pollard, C., Hartz, S., Leage, S.L., Paget, M.A., Cook, J., Enstone, A.: Elicitation of health state utilities associated with varying severities of flares in systemic lupus erythematosus. Health Qual. Life Outcomes **13**, 66 (2015). https://doi.org/10.1186/s12955-015-0262-0
17. Brazier, J., Ara, R., Azzabi, I., Busschbach, J., Chevrou-Severac, H., Crawford, B., Cruz, L., Karnon, J., Lloyd, A., Paisley, S., Pickard, A.S.: Identification, review, and use of health state utilities in cost-effectiveness models: an ISPOR Good practices for outcomes research task force report. Value Health **22**(3), 267–275 (2019). https://doi.org/10.1016/j.jval.2019.01.004
18. Xie, F., Zoratti, M., Chan, K., Husereau, D., Krahn, M., Levine, O., Clifford, T., Schunemann, H., Guyatt, G.: Toward a centralized, systematic approach to the identification, appraisal, and use of health state utility values for reimbursement decision making: introducing the health utility book (HUB). Med. Decis. Making **39**(4), 370–378 (2019). https://doi.org/10.1177/0272989x19837969
19. Walker, D.G., Wilson, R.F., Sharma, R., Bridges, J., Niessen, L., Bass, E.B., Frick, K.: Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. AHRQ Publication No. 12(13)-EHC132-EF. Agency for Healthcare Research and Quality (2012)
20. Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A.H., Mauskopf, J., Loder, E.: Consolidated health economic evaluation reporting standards (CHEERS)–explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value Health **16**(2), 231–250 (2013). https://doi.org/10.1016/j.jval.2013.02.002
21. Evers, S., Goossens, M., de Vet, H., van Tulder, M., Ament, A.: Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. Int. J. Technol. Assess. Health Care **21**(2), 240–245 (2005)
22. Nerich, V., Saing, S., Gamper, E.M., Holzner, B., Pivot, X., Viney, R., Kemmler, G.: Critical appraisal of health-state utility values used in breast cancer-related cost-utility analyses. Breast Cancer Res. Treat. (2017). https://doi.org/10.1007/s10549-017-4283-8
23. Chiou, C.F., Hay, J.W., Wallace, J.F., Bloom, B.S., Neumann, P.J., Sullivan, S.D., Yu, H.T., Keeler, E.B., Henning, J.M., Ofman, J.J.: Development and validation of a grading system for the quality of cost-effectiveness studies. Med. Care **41**(1), 32–44 (2003). https://doi.org/10.1097/01.mlr.0000039824.73620.e5
24. Simoens, S.: Assessment of methodological quality of economic evaluations in belgian drug reimbursement applications. PLoS One **8**(12), e85411 (2013). https://doi.org/10.1371/journal.pone.0085411
25. Stalmeier, P.F., Goldstein, M.K., Holmes, A.M., Lenert, L., Miyamoto, J., Stiggelbout, A.M., Torrance, G.W., Tsevat, J.: What should be reported in a methods section on utility assessment? Med. Decis. Making **21**(3), 200–207 (2001)
26. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., Gray, A.: Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. Pharmacoeconomics **33**(10), 985–991 (2015). https://doi.org/10.1007/s40273-015-0319-2
27. Xie, F., Pickard, A.S., Krabbe, P.F., Revicki, D., Viney, R., Devlin, N., Feeny, D.: A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). Pharmacoeconomics **33**(8), 867–877 (2015). https://doi.org/10.1007/s40273-015-0292-9
28. Brazier, J., Deverill, M., Green, C.: A review of the use of health status measures in economic evaluation. J. Health Serv. Res. Policy **4**(3), 174–184 (1999). https://doi.org/10.1177/135581969900400310
29. Drummond, M.F., Jefferson, T.O.: Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ **313**(7052), 275–283 (1996)
30. Ungar, W.J., Santos, M.T.: The pediatric quality appraisal questionnaire: an instrument for evaluation of the pediatric health economics literature. Value Health **6**(5), 584–594 (2003). https://doi.org/10.1046/j.1524-4733.2003.65253.x
31. Clemens, K., Townsend, R., Luscombe, F., Mauskopf, J., Osterhaus, J., Bobula, J.: Methodological and conduct principles for pharmacoeconomic research. Pharmaceutical Research and Manufacturers of America. PharmacoEconomics **8**(2), 169–174 (1995)
32. Adams, M.E., McCall, N.T., Gray, D.T., Orza, M.J., Chalmers, T.C.: Economic analysis in randomized control trials. Med. Care **30**(3), 231–243 (1992)
33. Gerard, K.: Cost-utility in practice: a policy maker's guide to the state of the art. Health Policy **21**(3), 249–279 (1992)
34. Sacristan, J.A., Soto, J., Galende, I.: Evaluation of pharmacoeconomic studies: utilization of a checklist. Ann. Pharmacother. **27**(9), 1126–1133 (1993). https://doi.org/10.1177/106002809302700919
35. Drummond, M., Manca, A., Sculpher, M.: Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. Int. J. Technol. Assess. Health Care **21**(2), 165–171 (2005)
36. Ramsey, S., Willke, R., Briggs, A., Brown, R., Buxton, M., Chawla, A., Cook, J., Glick, H., Liljas, B., Petitti, D., Reed, S.: Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. Value Health **8**(5), 521–533 (2005). https://doi.org/10.1111/j.1524-4733.2005.00045.x
37. Goetghebeur, M.M., Wagner, M., Khoury, H., Levitt, R.J., Erickson, L.J., Rindress, D.: Evidence and value: Impact on decision-Making–the EVIDEM framework and potential applications. BMC Health Serv. Res. **8**, 270 (2008). https://doi.org/10.1186/1472-6963-8-270
38. Davis, J.C., Robertson, M.C., Comans, T., Scuffham, P.A.: Guidelines for conducting and reporting economic evaluation of fall prevention strategies. Osteoporos Int. **22**(9), 2449–2459 (2011). https://doi.org/10.1007/s00198-010-1482-0
39. Vintzileos, A.M., Beazoglou, T.: Design, execution, interpretation, and reporting of economic evaluation studies in obstetrics. Am. J. Obstet. Gynecol. **191**(4), 1070–1076 (2004). https://doi.org/10.1016/j.ajog.2004.05.021
40. Grutters, J.P., Seferina, S.C., Tjan-Heijnen, V.C., van Kampen, R.J., Goettsch, W.G., Joore, M.A.: Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. Value Health **14**(5), 777–784 (2011). https://doi.org/10.1016/j.jval.2011.01.005
41. Russell, L.B., Gold, M.R., Siegel, J.E., Daniels, N., Weinstein, M.C.: The role of cost-effectiveness analysis in health and medicine. Panel on cost-effectiveness in health and medicine. Jama **276**(14), 1172–1177 (1996)
42. Siegel, J.E., Weinstein, M.C., Russell, L.B., Gold, M.R.: Recommendations for reporting cost-effectiveness analyses. Panel on cost-effectiveness in health and medicine. Jama **276**(16), 1339–1341 (1996)
43. Weinstein, M.C., Siegel, J.E., Gold, M.R., Kamlet, M.S., Russell, L.B.: Recommendations of the panel on cost-effectiveness in health and medicine. JAMA **276**(15), 1253–1258 (1996)
44. Sanders, G.D., Neumann, P.J., Basu, A., Brock, D.W., Feeny, D., Krahn, M., Kuntz, K.M., Meltzer, D.O., Owens, D.K., Prosser, L.A., Salomon, J.A., Sculpher, M.J., Trikalinos, T.A., Russell, L.B., Siegel, J.E., Ganiats, T.G.: Recommendations for conduct,

methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. JAMA **316**(10), 1093–1103 (2016). https://doi.org/10.1001/jama.2016.12195

45. Higgins, J.P.T., Altman, D.G., Gøtzsche, P.C., Jüni, P., Moher, D., Oxman, A.D., Savović, J., Schulz, K.F., Weeks, L., Sterne, J.A.C.: The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ **343** (2011). https://doi.org/10.1136/bmj.d5928
46. Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.: The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality Life Res. **19**(4), 539–549 (2010). https://doi.org/10.1007/s11136-010-9606-8

Authors and Affiliations

Michael J. Zoratti[1] · A. Simon Pickard[2] · Peep F. M. Stalmeier[3] · Daniel Ollendorf[4] · Andrew Lloyd[5] · Kelvin K W Chan[6,7] · Don Husereau[8] · John E. Brazier[9] · Murray Krahn[10] · Mitchell Levine[1,11] · Lehana Thabane[1,12] · Feng Xie[1] on behalf of Health Utility Book (HUB) Working Group

Michael J. Zoratti
zorattmj@mcmaster.ca

A. Simon Pickard
pickard1@uic.edu

Peep F. M. Stalmeier
Peep.Stalmeier@radboudumc.nl

Daniel Ollendorf
dollendorf@tuftsmedicalcenter.org

Andrew Lloyd
andrew.lloyd@acasterlloyd.com

Kelvin K W Chan
kelvin.chan@sunnybrook.ca

Don Husereau
don.husereau@gmail.com

John E. Brazier
j.e.brazier@sheffield.ac.uk

Murray Krahn
murray.krahn@theta.utoronto.ca

Mitchell Levine
levinem@mcmaster.ca

Lehana Thabane
thabanl@mcmaster.ca

[1] Department of Health Research Methodology, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[2] Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, The University of Illinois at Chicago, Chicago, IL, USA

[3] Health Evidence, Radboud University Medical Centre, Nijmegen, The Netherlands

[4] Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

[5] Acaster Lloyd Consulting Ltd, London, UK

[6] Sunnybrook Odette Cancer Centre, Toronto, ON, Canada

[7] Canadian Centre for Applied Research in Cancer Control, Toronto, Canada

[8] School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

[9] School of Health and Related Research, University of Sheffield, Sheffield, UK

[10] Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada

[11] Department of Medicine, Division of Clinical Pharmacology and Toxicology, McMaster University, Hamilton, ON, Canada

[12] St Joseph's Healthcare, Hamilton, ON, Canada