IDENTIFYING AND APPLYING HEALTH UTILITIES IN DECISION-MAKING

TOWARDS TRANSPARENCY IN THE SELECTION, APPRAISAL, AND APPLICATION OF HEALTH STATE UTILITY VALUES IN DECISION-MAKING

By MICHAEL JAMES ZORATTI, B.Sc., M.Sc.

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

Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2021) Health Research Methodology	McMaster University Hamilton, Ontario
Title	Towards transparency in the selection, appraisal, and application of health state utility values in decision-making
Author	Michael J Zoratti B.Sc. University of Ottawa, 2012 M.Sc. McMaster University, 2014
Supervisor	Dr. Feng Xie
Committee	Dr. Mitchell Levine Dr. Lehana Thabane
Number of pages	xiv, 174

LAY ABSTRACT

This doctoral thesis was designed to explore existing methodologies and propose novel approaches to improve the comprehensiveness, relevance, and transparency in the development of health economic evaluations. A systematic literature review protocol, as well as a sample of completed work, will support investigators and decision-makers in identifying the available published health utility literature when building or critiquing health economic models. A methods-focused review follows, which highlights an important shortcoming in the tools available to researchers in the evaluation of both the methodological rigour and goodness-of-fit of health utility inputs in their decision contexts. The Health utility Application Tool is then proposed as a solution to the latter issue. It is intended to assist investigators in communicating and defending their reasonings when selecting among the health utility literature. Overall, this body of research is intended as a simple but practical solution to a persistent problem in the health economist's toolbox.

ABSTRACT

The identification and selection of health utility inputs for health economic models is an essential yet often overlooked task in the development of useful, context-relevant models. This doctoral thesis was designed to explore existing methodologies and propose novel approaches to improve the comprehensiveness, relevance, and transparency of these processes.

A broadly scoped systematic literature review was designed to identify published health utility estimates in the field of oncology. The objective of this work is to support both researchers and, importantly, reviewers who may not have sufficient resources available to evaluate the available literature and determine whether other viable inputs may have been suitable to inform a model parameter. This review infrastructure will gradually be expanded across the clinical spectrum.

To evaluate the current framing of health utility-relevant concepts in critical appraisal and reporting checklists, a methods-focused literature review was conducted. While several health economic evaluation- and health utility-relevant tools and checklists were identified, this review confirmed that none were designed for the express purpose of systematically critically appraising health utility studies or evaluating context relevance.

Finally, the Health utility Application Tool (HAT) was developed based on the learnings from previous research activities and engagement with knowledge users and doers in the field of health technology assessment. This tool is intended to guide researchers in selecting the most relevant evidence for models, with several questions directed toward clinical and jurisdiction-specific context. Framed primarily as a

communication tool, the HAT is intended for researchers to use to document reasoning rather than for critically appraising others' work.

This body of research has established the groundwork upon which further developments may emerge. The success of these efforts will be realized over time and necessitates a commitment to knowledge translation and dissemination with sustained engagement across the research community.

ACKNOWLEDGEMENTS

This work reflects years of collaboration and commitment that would have not been possible without the ongoing support and encouragement from so many.

A very sincere thank you to my doctoral supervisor, Dr. Feng Xie. Your generosity with your time and thoughts and your unwavering support for me to explore new challenges and opportunities has made me a better researcher. It has been a pleasure to study under you and I look forward to continuing to pursue further research endeavours in the future. I am forever grateful for your tutelage.

My thesis committee members, Dr. Mitchell Levine and Dr. Lehana Thabane, have been instrumental to my success and confidence as I prepare the culmination of my doctoral research. You have both been generously available for advice and guidance throughout my graduate studies, be it through committee meetings, independent coursework, the comprehensive examination independent study, or for general counsel about the offerings of the health research methodology program.

Thank you to the many mentors I have been blessed with through my graduate studies. The advice and support I received shaped my growth in the field of health methods research and positioned me to where I am today. Thank you also to the research group who supported my projects, both those which comprise this thesis and the others we have engaged in. I look forward to continuing our collaborations as we continue to build off of this body of research. And, of course, a very big thank you to the administrative team in the Health research methodology, Evidence, and Impact office. I can't imagine where any of us would be without your dedication and commitment to guiding us all to excellence.

To my fellow graduate students, thank you for your friendship and collaborations throughout my time at McMaster University. I have learned so much from each of you, whether through classes, projects, or casual conversations in the student room. I look forward to seeing where your careers take you.

And, finally, thank you to my family. Thank you to my wife, Emily, for consistently supporting my ambitions, celebrating our achievements, and for being eternally dedicated to us. Thank you to my parents for your steadfast encouragement to pursue this education. And thank you to my brother and my sister-in-law for being a consistent source of positivity. This thesis is dedicated to my daughters, Georgina and Violet. May you always be curious about the world and be relentless in the pursuit of your dreams.

TABLE OF CONTENTS

Lay abstract iii
Abstract iv
Acknowledgements vi
Table of contents vii
List of tables x
List of figures and boxes xi
List of abbreviations xii
Declaration of academic achievement xiii
Chapter 1 – Background, Motivation, and Purpose 1
A brief note on advances in health research 1
Concept of health utility 2
Applications of health utility 2
Are all health utilities created equal? 3
Whose preferences should be considered? 4
Health utilities in the context of Canadian health technology assessment 6
Sections of this thesis 8
References 10
Chapter 2-1 – A protocol for a systematic literature review of health state utility values in oncology 15
Abstract 17
Background 19
Objectives 21
Methods 21
Pilot review 26
Discussion 28
References 30
Tables and Figures 40
Appendix A 48

Chapter 2-2 – A systematic literature review of health state utility values in bladder cancer 51
Abstract 53
Background 55
Methods 56
Results 57
Discussion 61
Conclusion 64
References 65
Tables and Figures 68
Chapter 3 – Evaluating the conduct and application of health utility studies: A review of critical appraisal tools and reporting checklists 74
Abstract 76
Introduction 78
Methods 81
Results 82
Discussion 86
Conclusion 89
References 90
Tables and Figures 96
Appendix A 106
Chapter 4 – Towards transparency in the selection of published health utility inputs in cost-utility analyses: The Health utility Application Tool (HAT) 128
Acknowledgements 130
Abstract 131
Introduction 133
Methods 134
Results 136
Discussion 142

Conclusion 146
References 147
Tables and Boxes 149
Appendix A: Telephone interview script 155
Appendix B: Online survey questionnaire 162
Chapter 5 – Conclusions 171

LIST OF TABLES

Chapter 2-1

Table 1: Search strategy for MEDLINE and EMBASE (via Ovid)..... 41
Table 2: Characteristics of included studies of health utility in thyroid cancer..... 43
Table 3: Health states in thyroid cancer..... 44

Chapter 2-2

Table 1: Study and respondent characteristics of health utility studies in bladder cancer
..... 69
Table 2: Health states and utility values reported for bladder cancer..... 70

Chapter 3

Table 1: Summary of 12 tools included in the review..... 97
Table 2: Items of relevance for assessing quality and applicability of health utility studies
..... 100

Chapter 4

Table 1: Interview and online survey participants..... 150
Table 2: Results of thematic analysis..... 151

LIST OF FIGURES AND BOXES

Chapter 2-1

Figure 1: PRISMA flow of information diagram 47

Chapter 2-2

Figure 1: Summary of the study selection process 72
Figure 2: Health utility estimates for the bladder cancer health state 73

Chapter 3

Figure 1: Process of selecting tools and relevant items 105

Chapter 4

Box 1: The Health utility Application Tool (HAT) 153

LIST OF ABBREVIATIONS

Abbreviation	Definition
AHRQ	Agency for Healthcare Research and Quality
BMJ	British Medical Journal
CADTH	Canadian Agency for Drugs and Technologies in Health
CEA	Cost-effectiveness analysis
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CI	Confidence interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COSMIN	COnsensus-based Standards for the selection of health status Measurement INstruments
CREATE	Checklist for REporting VAluaTion StudiEs
CUA	Cost utility analysis
EMBASE	Excerpta Medica dataBASE
HAT	Health utility Application Tool
HRQoL	Health-related quality of life
HSU	Health state utility
HSUV	Health state utility value
HTA	Health technology assessment
HUB	Health Utility Book
HUI	Health Utility Instrument
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MAPS	MApping onto Preference-based measures reporting Standards
MAUI	Multi-attribute utility instrument
MEDLINE	Medical Literature Analysis and Retrieval System Online
MEPS	Medical Expenditure Panel Survey
NICE	National Institute for Health and Care Excellence
PICO	Population, Intervention, Comparator, Outcome
PQAQ	Pediatric Quality Appraisal Questionnaire
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
QALY	Quality-adjusted life year
QHES	Quality of Health Economic Studies
SEER	Surveillance, Epidemiology, and End Results
SEM	Standard error of the mean
SG	Standard gamble
TTO	Time trade-off
VAS	Visual analogue scale

DECLARATION OF ACADEMIC ACHIEVEMENT

This sandwich thesis consists of four manuscripts. At the time of writing, three (Chapter 2-1; Chapter 3; Chapter 4) have been submitted, accepted, and published in peer reviewed journals. The version of Chapter 2-1 presented in this thesis reflects minor revisions based on a re-evaluation of the eligibility of the publications included in the pilot review work. These revisions have not impacted the integrity or conclusions of this manuscript. The fourth manuscript (Chapter 2-2) is presented here but not yet submitted for publication.

Michael Zoratti led the research activities for each manuscript, in collaboration with his doctoral supervisor Dr. Feng Xie, and was the primary researcher responsible for conceptualizing the study designs and protocols, obtaining ethics approval as applicable, preparing study materials, conducting data gathering exercises, summarizing the evidence, drafting manuscripts and soliciting feedback from peers, coordinating submissions to journals for peer review, and preparing responses to reviewer comments. The co-authors named for each manuscript contributed to the conceptual study design, interpretation of findings, evaluation of scientific integrity, and preparation of dissemination materials.

This work was conducted between September 2016 and July 2021. Author affiliations listed for each of the included manuscripts are based on affiliations at the time of writing or publishing.

Chapter 1 – Introduction
This chapter is unpublished. MZ is the sole author.

Chapter 2-1 – A protocol for a systematic literature review of health state utility values in oncology
This chapter was published in Medical Decision Making: Policy & Practice on August 13, 2019. MZ and FX conceived the study design. MZ led the research activities with support from TZ and FX. MZ prepared the initial manuscript draft. All authors critically reviewed the manuscript for scientific integrity and approved its submission for publication.

Chapter 2-2 – A systematic literature review of health state utility values in bladder cancer
This chapter is unpublished. MZ led the research activities based on Chapter 2-1.

Chapter 3 – Evaluating the conduct and application of health utility studies: A review of critical appraisal tools and reporting checklists
This chapter was published in the European Journal of Health Economics on April 11, 2021. It has been reproduced here with permission (License 5113680379037). MZ and FX conceived the study design. MZ led the research activities with support from FX. MZ prepared the initial manuscript draft. All authors critically reviewed the manuscript for scientific integrity and approved its submission for publication.

Chapter 4 – Towards transparency in the selection of published health utility inputs in cost-utility analyses: The Health utility Application Tool (HAT)
This chapter was published in PharmacoEconomics on June 21, 2021. It has been reproduced here with permission (License 5113680733458). All authors conceived the study concept. MZ was responsible for leading the study design and execution, including preparing data collection materials and obtaining ethics approval. MZ and FX collected and conducted an initial interpretation of the data. MZ and FX drafted the initial manuscript. All authors critically reviewed the manuscript for scientific integrity and approved its submission for publication.

Chapter 5 – Conclusions
This chapter is unpublished. MZ is the sole author.

CHAPTER 1 – BACKGROUND, MOTIVATION, AND PURPOSE

A brief note on advances in health research

Investigators continue to challenge precedent and pursue novel approaches to evaluating and explaining clinical phenomena, furthering the incredible methodological research advancements witnessed over recent decades. This evolution has been aided by technological breakthroughs which have transformed the way research is conducted, from the rapid retrieval of literature to the evaluation of thousands of simulations to the dissemination of study findings around the globe.

As these methodological evolutions have continued to thrive, so too has the toolbox to evaluate them. Today, nearly every health outcome research endeavour may be accompanied by some appraisal or checklist to guide it. These vary considerably in scope and purpose, varying from reporting checklists, which encourage investigators to provide readers with sufficient information to evaluate the methods and results, to critical appraisal tools designed to evaluate the methodological rigour with which a study was conducted. These approaches need not be contrasted against one another, but rather should be positioned as complementary views on establishing validity. While the merits of specific tools and approaches may be debated – a full discussion of these points is outside the scope here – it may be agreed that, overall, they offer more benefit than harm in promoting good science and standards.

Yet, while the development of methods has been widespread, it has not been equal. This thesis was designed to address perceived methodological shortcomings in the evaluation and application of the published health utility literature to cost utility analyses.

Concept of health utility

Health utilities are cardinal values which reflect preferences for health states. The anchors for this metric are 0 for dead and 1 for perfect health, though negative values may represent states of health considered worse than dead. In these cases, the lower bound is dependent on the scaling measure used.

Several methodologies have evolved to facilitate the estimation of health utilities. Broadly, these may be labelled as either direct or indirect methods.[1] Direct measures include rating scales, such as the visual analogue scale (VAS), or willingness to trade-off methods, such as the standard gamble (SG) or time trade-off (TTO). However, the implementation of these measures is time consuming and may be prohibitively burdensome for many applications. In response, indirect methods have been developed, including both generic and condition-specific preference-based health status or health-related quality of life (HRQoL) instruments.[2,3] Also labelled multi-attribute utility instruments (MAUIs), these tools have been well received by the research community for their ease of implementation as they require minimal oversight or direction by the investigator. The responses to the questions in these instruments can be converted into health utility values by applying a predeveloped tariff or value set, a catalogue of values for all possible health states an instrument can describe. Importantly, there is significant evidence emerging in the literature to suggest that the selection of a relevant value set has an important impact on the health utility estimates which one may generate.[4-7]

Applications of health utility

Health utility estimates are often associated with their most common application, the cost utility analysis (CUA). This form of economic evaluation, a type of cost effectiveness analysis (CEA), measures health outcomes using the quality-adjusted life year (QALY). The QALY is a more holistic outcome than the natural, and often disease-specific, clinical outcome units used in a traditional CEA. Decision makers across many jurisdictions globally recommend the use of CUA as a major component of reimbursement dossiers given the outputs are readily interpretable and portable across clinical contexts.

Are all health utilities created equal?

There are several established approaches to eliciting preferences and utility estimates. Yet, while they may be implemented with equal methodological rigour, they do not necessarily return identical results.[8-18] For example, a review in prostate cancer found that the difference in utility estimates exceeded 0.5 for three states and were between 0.1 and 0.5 for eight states.[9] This observation is not an anomaly, with large variances in estimates frequently observed in many diseases such as cancer,[9,13,14,17] osteoporosis,[19] surgical site infection,[20] cataracts,[16] haemophilia,[21] liver disease,[22] chronic obstructive pulmonary disease,[23] and inflammatory bowel disease.[24] And while there are several approaches to collecting this evidence, methodological shortcomings do persist in the literature.[9,11,19,25-27] For instance, a review by Sturza and colleagues identified 223 unique health utility inputs in lung cancer across 23 studies.[15] Interestingly, approximately 16% of these estimates were based on judgment, rather than an established, validated

technique. Thus, there arises a need to be cognizant of the selections one makes when sourcing evidence for an economic evaluation.

These considerations become particularly important in the context of reimbursement decision making where these selections may impact the value messaging of a novel therapeutic and, ultimately, its ability to be accessed by patients. Yet, a 2011 review by Tosh and colleagues of 71 National Institute for Health and Clinical Excellence (NICE) technology appraisals revealed that only 31% of the 39 submissions which based health utility inputs on published literature identified this evidence through a systematic literature review.[28] The absence of a systematic approach to identifying published health utility evidence may increase the risk of bias while also underestimating the uncertainty of both the health utility estimates for various health states and, consequently, the cost-effectiveness estimates of the CUAs in which they are applied. It is important to note that these observations are not to say that these selections were made with the intent of distorting or obscuring the evidence. Both investigators who make such selections and decision makers who assess their merits in the given reimbursement question engage in this activity with an intent to appropriately reflect the decision context. However, evaluations by both parties necessitate careful consideration of the evidence base as well as a means of processing the strengths and weaknesses of the available literature.

Whose preferences should be considered?

The debate on whose preferences should be used persists.[29] Through a systematic literature review, Helgesson and colleagues concluded that respondents with experience of the health state are expected to generate the most accurate information.[30] Several

themes emerged from their evaluation of the literature, summarized as relating to theoretical arguments, the effects of valuing, and failures of the general public to value health states. The latter argument posits a superior patient knowledge, the ability for individuals to adapt to their circumstances, and valuation difficulties relating to reference points and valuation compression. Indeed, experience with the health state being evaluated tends to be the most significant differentiator between patients and the general public.

Proxy respondents, such as family members or caregivers, are a typical substitute in contexts where patients are unable to complete evaluations themselves due to age or cognitive impairment. In a 2018 systematic review and meta-analysis of childhood health utilities, Kwon and colleagues found that health utility estimates from parental proxies tended to overestimate children's HRQoL compared to responses provided directly by children.[31] Systematic differences between children's self-reported health status and the estimates from proxy respondents,[32-34] as well as between different types of proxies,[35-37] have been regularly observed. However, investigators are left with few options for evaluating HRQoL in certain populations.

Systematic differences in observed outcomes across respondent types should be considered and acknowledged by analysts conducting health economic evaluations. There are philosophical arguments which may be made for either camp.[29] For example, one may recommend that outcomes from the informed general public be used when the economic evaluation is intended to guide funding allocations in a publicly-funded healthcare system. This algins with the principle of justice as respondents from the general public

are shrouded by the 'veil of ignorance' which allows them to maintain impartiality and lack bias in their evaluation of the health state. Yet this runs in contravention to the welfarist approach which argues that individuals are the best judges of their own welfare. Another important consideration is the consistency in methodology when choosing model inputs in a CUA. Where this is not feasible, investigators should identify and acknowledge possible limitations in the ability to accurately value the model's health outcomes.

Health utilities in the context of Canadian health technology assessment

Formal guidance provides select considerations for investigators to weigh in the application of health utility estimates. As the preeminent authority on health technology assessment (HTA) methodologies in Canada, the Canadian Agency for Drugs and Technologies in Health (CADTH) issues regular guidance and recommendations for the conduct of economic evaluations. CADTH maintains that a CUA should be the reference case economic evaluation. There are several references to health utility throughout the 4th edition of the CADTH 'Guidelines for the Economic Evaluation of Health Technologies':[38]

> "Factors that may lead to different estimates of costs and outcomes associated with interventions across distinct subgroups of the population should be specified. These could be factors that affect the natural history of disease, the effectiveness of treatments, or the utilities or costs associated with the disease or treatments."
>
> "Health preferences (i.e., utilities) should reflect the health states in the model and be conceptualized to address the decision problem."
>
> "10.5 The selection of data sources for health state utility values should be based on their fitness for purpose, credibility, and consistency. Describe the trade-offs among these criteria and provide justification for the selected sources."

"Based on their ease of use, comparability, and interpretability, it is recommended that, in the reference case, researchers use utilities from an indirect method of measurement that is based on a generic classification system. The selection of a particular indirect generic method should be based on its fitness in terms of reflecting the health states of interest and their associated attributes, and in terms of capturing potentially important changes within and among states."

"The researcher should also assess the credibility of the instrument in terms of whether it represents an established instrument with demonstrated psychometric properties including feasibility, reliability, and validity. Consistency with respect to the data used for estimating utility values is strongly recommended; in particular, it is recommended that utility data from the same instrument and population be used to estimate all the utilities in an economic evaluation."

"When considering Canadian-specific preference sets, researchers must weigh their fitness for purpose against any issues of credibility and consistency with the data used to inform other parameter estimates, to ensure that the specific instrument addresses the decision problem."

"There may be instances where utilities for the applicable health states have already been estimated based on indirect generic methods of measurement. These may have been gathered as part of routine data sets, published in the literature, or collected alongside a study. In such cases, these utilities should be critically assessed according to the previously discussed criteria of fitness for purpose, credibility, and consistency, to ensure that they reflect the health states of interest in the model and that the preferences reflect those of the general population."

"When searching for utilities, it is important that the search methods are comprehensive and presented in a transparent manner such that they can be replicated by others (i.e., similar to the approach for searching for data on clinical effects, as detailed in the Effectiveness section)."

"Similar to the selection of other data inputs, when selecting utility estimates from among multiple indirect generic measures, researchers should employ judgment regarding which data source represents the best estimate, weighing trade-offs among the criteria of fitness for purpose, credibility, and consistency. Based on a consideration of these criteria, the researcher must justify the selection of the utilities used in the reference case analysis. Regardless of the data source(s) used to inform utility estimates, probability distributions for each utility value should be derived and the associated uncertainty propagated through the model. The potential implications of any trade-offs among utility estimates should be considered in the context of the probabilistic analysis or using scenario analysis."

> "In the context of the social decision-making viewpoint adopted in these Guidelines, the valuation of health state utilities should be based on the preferences of the general population."
>
> "Justification for the use of preferences other than those of the general population, as well as the methods for measuring and valuing the utilities should be provided and clearly described so that the implications can be assessed relative to the reference case results. This would be particularly important when outcomes of the analysis are sensitive to preference weights."
>
> "Economic models may involve states of health defined by a combination of health states (i.e., joint health states). Ideally, utilities may be obtained for these joint health states. However, it is often not possible to identify utilities that fully reflect the combination of health states."

Notably, there are repeated calls for the inputs to economic evaluations, including health utility values, to be "critically assessed" for "criteria of fitness for purpose, credibility, and consistency." While these statements encourage good practices, investigators may not have the tools to systematically evaluate this fit, both in terms of specific criteria and means. Thus, there is an opportunity to provide investigators with a formal approach which is grounded in good methods and tailored to the needs of both modellers charged with defending selections and the decision makers tasked with assessing those rationalizations. This opportunity motivates this thesis.

Sections of this thesis

This first chapter lays the groundwork for this thesis by introducing key concepts and their interpretation and application in the field of health utility research and, broadly, economic evaluations. The second chapter introduces the systematic literature review methodology for the Health Utility Book (HUB) registry for health utility estimates in oncology. The second part of chapter two presents a sample output from this registry. The third chapter summarizes the available literature for the evaluation of health utility inputs

in economic evaluations. This is achieved through a literature review of published scales and checklists, with consideration of the means by which they were developed, their intended uses, and how concepts of health utility are reflected. This chapter concludes that the economic evaluation appraisal toolbox is lacking with respect to its ability to provide sufficient guidance on the evaluation of health utility study methodological rigour and goodness of fit. From this follows the fourth chapter, which presents the development of Health utility Application Tool (HAT). The HAT is an answer to the lack of concrete direction on appraising the goodness of fit of potential health utility inputs for an economic model. Finally, the merits of the collective works of this thesis are summarized and framed in the context of what is needed to have made them worthwhile. Additional investigations are highlighted which build from this work and their grounding in the promotion of good research methodology is emphasized.

References

1. Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2006;9(4):213-218.
2. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess.* 2014;18(9):1-224.
3. Goodwin E, Green C. A Systematic Review of the Literature on the Development of Condition-Specific Preference-Based Measures of Health. *Applied health economics and health policy.* 2016;14(2):161-183.
4. Zhao Y, Li S-P, Liu L, Zhang J-L, Chen G. Does the choice of tariff matter?: A comparison of EQ-5D-5L utility scores using Chinese, UK, and Japanese tariffs on patients with psoriasis vulgaris in Central South China. *Medicine (Baltimore).* 2017;96(34):e7840-e7840.
5. Wang H, Cao C, Guo C, et al. An evaluation of EQ-5D-3L health utility scores using five country-specific tariffs in a rural population aged 45–69 years in Hua county, Henan province, China. *Health and quality of life outcomes.* 2020;18(1):228.
6. Kiadaliri AA, Eliasson B, Gerdtham U-G. Does the choice of EQ-5D tariff matter? A comparison of the Swedish EQ-5D-3L index score with UK, US, Germany and Denmark among type 2 diabetes patients. *Health and quality of life outcomes.* 2015;13(1):145.
7. Bosmans JE, Rossenaar MM, van Dongen JM, van Tulder MW. The Impact of Country-Specific Utility Tariffs on the Outcome of Cost-Utility Analyses; A Case Study. *Value in Health.* 2016;19(7):A386.
8. Smith-Palmer J, Bae JP, Boye KS, Norrbacka K, Hunt B, Valentine WJ. Evaluating health-related quality of life in type 1 diabetes: a systematic literature

review of utilities for adults with type 1 diabetes. *ClinicoEconomics and outcomes research : CEOR.* 2016;8:559-571.

9. Bremner KE, Chong CA, Tomlinson G, Alibhai SM, Krahn MD. A review and meta-analysis of prostate cancer utilities. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2007;27(3):288-298.
10. Carter GC, King DT, Hess LM, et al. Health state utility values associated with advanced gastric, oesophageal, or gastro-oesophageal junction adenocarcinoma: a systematic review. *Journal of medical economics.* 2015;18(11):954-966.
11. Djalalov S, Rabeneck L, Tomlinson G, Bremner KE, Hilsden R, Hoch JS. A Review and Meta-analysis of Colorectal Cancer Utilities. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2014;34(6):809-818.
12. Hao Y, Wolfram V, Cook J. A structured review of health utility measures and elicitation in advanced/metastatic breast cancer. *ClinicoEconomics and outcomes research : CEOR.* 2016;8:293-303.
13. Paracha N, Thuresson PO, Moreno SG, MacGilchrist KS. Health state utility values in locally advanced and metastatic breast cancer by treatment line: a systematic review. *Expert review of pharmacoeconomics & outcomes research.* 2016;16(5):549-559.
14. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. *Expert review of pharmacoeconomics & outcomes research.* 2010;10(5):553-566.
15. Sturza J. A review and meta-analysis of utility values for lung cancer. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2010;30(6):685-693.
16. Hahn U, Krummenauer F. Results and methodology of cost-utility evaluation of cataract surgery in developed countries: Quality-adjusted life years and cataract. *J Cataract Refract Surg.* 2017;43(6):839-847.

17. Peasgood T, Herrmann K, Kanis JA, Brazier JE. An updated systematic review of Health State Utility Values for osteoporosis related conditions. *Osteoporos Int.* 2009;20(6):853-868.
18. Nerich V, Saing S, Gamper EM, et al. Critical appraisal of health-state utility values used in breast cancer-related cost-utility analyses. *Breast cancer research and treatment.* 2017;164(3):527-536.
19. Brazier JE, Green C, Kanis JA. A systematic review of health state utility values for osteoporosis-related conditions. *Osteoporos Int.* 2002;13(10):768-776.
20. Gheorghe A, Moran G, Duffy H, Roberts T, Pinkney T, Calvert M. Health Utility Values Associated with Surgical Site Infection: A Systematic Review. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2015;18(8):1126-1137.
21. Grosse SD, Chaugule SS, Hay JW. Estimates of utility weights in hemophilia: implications for cost-utility analysis of clotting factor prophylaxis. *Expert review of pharmacoeconomics & outcomes research.* 2015;15(2):267-283.
22. McLernon DJ, Dillon J, Donnan PT. Health-state utilities in liver disease: a systematic review. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2008;28(4):582-592.
23. Moayeri F, Hsueh YS, Clarke P, Hua X, Dunt D. Health State Utility Value in Chronic Obstructive Pulmonary Disease (COPD); The Challenge of Heterogeneity: A Systematic Review and Meta-Analysis. *Copd.* 2016;13(3):380-398.
24. Malinowski KP, Kawalec P. Health utility of patients with Crohn's disease and ulcerative colitis: a systematic review and meta-analysis. *Expert review of pharmacoeconomics & outcomes research.* 2016;16(4):441-453.
25. Campolina AG, Rozman LM, Decimoni TC, Leandro R, Novaes HM, De Soárez PC. Many Miles to Go: A Systematic Review of the State of Cost-Utility Analyses in Brazil. *Applied health economics and health policy.* 2017;15(2):163-172.

26. Nerich V, Saing S, Gamper EM, et al. Cost-utility analyses of drug therapies in breast cancer: a systematic review. *Breast cancer research and treatment.* 2016;159(3):407-424.
27. Doyle S, Lloyd A, Moore L, Ray J, Gray A. A systematic review and critical assessment of health state utilities: weight change and type 2 diabetes mellitus. *PharmacoEconomics.* 2012;30(12):1133-1143.
28. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14(1):102-109.
29. Stamuli E. Health outcomes in economic evaluation: who should value health? *British Medical Bulletin.* 2011;97(1):197-210.
30. Helgesson G, Ernstsson O, Åström M, Burström K. Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. *Qual Life Res.* 2020;29(6):1465-1482.
31. Kwon J, Kim SW, Ungar WJ, Tsiplova K, Madan J, Petrou S. A Systematic Review and Meta-analysis of Childhood Health Utilities. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2018;38(3):277-305.
32. Brunner HI, Klein-Gitelman MS, Miller MJ, et al. Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum.* 2004;51(5):763-773.
33. Kulpeng W, Sornsrivichai V, Chongsuvivatwong V, et al. Variation of health-related quality of life assessed by caregivers and patients affected by severe childhood infections. *BMC Pediatrics.* 2013;13(1):122.
34. Wolke D, Chernova J, Eryigit-Madzwamuse S, Samara M, Zwierzynska K, Petrou S. Self and parent perspectives on health-related quality of life of adolescents born very preterm. *J Pediatr.* 2013;163(4):1020-1026.e1022.

35. Kramer MS, Etezadi-Amoli J, Ciampi A, et al. Parents' versus physicians' values for clinical outcomes in young febrile children. *Pediatrics.* 1994;93(5):697-702.
36. Janse AJ, Uiterwaal CS, Gemke RJ, Kimpen JL, Sinnema G. A difference in perception of quality of life in chronically ill children was found between parents and pediatricians. *J Clin Epidemiol.* 2005;58(5):495-502.
37. Saigal S, Stoskopf BL, Feeny D, et al. Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. *Jama.* 1999;281(21):1991-1997.
38. CADTH. *Guidelines for the economic evaluation of health technologies: Canada. 4th ed.* Ottawa: CADTH; March 2017 2017.

CHAPTER 2-1 – A PROTOCOL FOR A SYSTEMATIC LITERATURE REVIEW OF HEALTH STATE UTILITY VALUES IN ONCOLOGY

Status: This manuscript was published in Medical Decision Making: Policy and Practice on August 13, 2019. The version presented in this thesis reflects minor revisions based on a re-evaluation of the eligibility of the publications included in the pilot review work. These revisions have not impacted the integrity or conclusions of this manuscript.

Original publication:

Zoratti MJ, Zhou T, Chan K, et al. Health Utility Book (HUB)–Cancer: Protocol for a Systematic Literature Review of Health State Utility Values in Cancer. *MDM Policy & Practice.* 2019;4(2):2381468319852594.

A protocol for a systematic literature review of health state utility values in oncology

Michael J Zoratti,[1] Ting Zhou,[1,2] Kelvin Chan,[3] Oren Levine,[4] Murray Krahn,[5] Don Husereau,[6] Tammy Clifford,[7] Holger Schünemann,[1] Gordon Guyatt,[1] Feng Xie[1]

Author affiliations:

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

[2]School of International Pharmaceutical Business, China Pharmaceutical University, Nanjing, China

[3]Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

[4]Department of Oncology, McMaster University, Hamilton, Ontario, Canada

[5]Toronto General Hospital, Toronto, Ontario, Canada

[6]Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada

[7]Canadian Institutes for Health Research, Ottawa, Ontario, Canada

Abstract

Background: Treatment options in oncology are rapidly advancing and public payer systems are increasing under pressure to adopt new but expensive cancer treatments. Cost-utility analyses (CUAs) are used to estimate the relative costs and effects of competing interventions, where health outcomes are measured using quality-adjusted life years (QALYs). Health state utility values (HSUVs) are used to reflect health-related quality of life or health status for the calculation of QALY. To support reimbursement agencies in the appraisal of oncology drug submissions, which typically include a CUA component, we have proposed a systematic literature review of published HSUV estimates in the field of oncology.

Methods: The following databases will be searched: MEDLINE, EMBASE, EconLit, and CINAHL. A team of reviewers, working independently and in duplicate, will evaluate abstracts and full-text publications for eligibility against broad inclusion criteria. Studies using a direct, indirect, or combination approach to eliciting preferences related to cancer or cancer treatments are eligible. Data extraction will capture details of study methodology, participants, health states, and corresponding HSUVs. We will summarize our findings with descriptive analyses at this stage. A pilot review in thyroid cancer is presented to illustrate the proposed methods.

Discussion: This systematic review will generate a comprehensive summary of the oncology HSUV literature. As a component of the Health Utility Book (HUB) project, we anticipate that this work will assist both health economic modellers as well as critical reviewers in the development and appraisal of CUAs in oncology.

Registration: CRD42018095049

Background

Reimbursement practices have evolved over time. Health technology assessment (HTA) represents a comprehensive approach to the evaluation of emerging and existing health care interventions. A major component of HTA submissions to public-payer drug plans are economic evaluations designed to compare competing interventions with respect to both clinical and economic consequences. Cost utility analyses (CUAs) are a type of economic evaluation accepted by several major HTA bodies worldwide.[1-3] These analyses are particularly useful for interventions or programs that not just extend life but also impact the patient's health-related quality of life or health status. In a CUA, health state utility values (HSUVs) are used to estimate quality-adjusted life years (QALYs) which provide an estimate of both quantity and quality of life [4]. Health states may be simple or complex, defined by several factors, including disease and treatment characteristics as well as functionality and limitations. By convention, full or perfect health is assigned the maximum utility score of 1, while death is assigned a utility score of 0. Health states with negative utilities are perceived as worse than death. While HSUV estimates are inherently subjective, the general rule is to weight more preferable health states with higher utility.

Investigators can measure HSUVs using different approaches. These can be through direct measurements, such as the visual analogue scale (VAS), or through preferences elicited using probabilities or willingness to trade-off between quantity and quality of life methods, such as the standard gamble (SG) or time trade-off (TTO) technique. Indirect methods using multi-attribute utility-based instruments (MAUIs), relatively short questionnaires that require only a few minutes to complete, provide an

easier alternative to these direct measures. The term indirect measure is used as individual patients do not explicitly provide preferences for their health states but rather describe their state which is then assigned a value according to a scoring algorithm which has been developed separately with the general public or patient groups. Investigators have used both direct and indirect methods in addressing HSUV for cancer treatments.[5]

Rapid progress in the field of oncology has given way to new therapies, and these treatments have made significant contributions to prolonging life expectancies or improving quality of life.[6] However, these new cancer treatments often come at high costs.[7,8] As part of the drug reimbursement process in many countries, manufacturers are required to submit dossiers containing CUAs. It is recommended that HSUV inputs for these CUAs be identified through systematic literature reviews.[9] However, a cross-sectional review of 71 technology appraisals submitted to the National Institute for Health and Clinical Excellence (NICE) found that out of 39 submissions that obtained HSUVs from published studies, only 31% adopted a systematic approach.[10] Moreover, even when these reviews are commissioned, the review's methods and findings may not be made explicit to reimbursement agencies. Thus, the complete evidence profile for a given health state may not be clear to reviewers and their ability to make a critical appraisal of model inputs may be limited.

There are several systematic reviews of HSUVs across the cancer spectrum.[11-15] However, these reviews are targeted to answer a question in a particular area of oncology. To date, no central catalogue of cancer-related HSUVs has been established. Such a resource would dramatically enhance researchers' ability to select and evaluate the

available health utility literature in a thorough and timely manner. In the interest of the comprehensive, accountable, and transparent evaluation of CUAs in the support of oncology drug reimbursement submissions, we have developed a systematic review protocol to identify and describe published health states and HSUVs across the spectrum of oncology research. This review is part of the Health Utility Book (HUB) as described by Xie et al.[16]

Objectives

The objective of this paper is to present a study protocol for systematically identifying and describing the health utility literature in the field of oncology with respect to both the methods used and the estimates attained for cancer-related health states and a pilot study applying this protocol to thyroid cancer.

Methods

This systematic review will be conducted in general accordance with published guidelines and good practices.[17-22] Additional considerations have been made given the broad scope of the review and the anticipated volume of work. The traditional Patient, Intervention, Comparison, Outcome (PICO) statement, common to clinical research, is not typically amenable to reviews of HSUVs.[17,18] Specifically, this review is not designed to collect information on any particular intervention or comparator. This protocol has been registered with the Prospective Register of Systematic Reviews (PROSPERO CRD42018095049).

Search strategy

The scope of the search strategies for this review is purposefully broad. The databases to be queried include: MEDLINE via Ovid; EMBASE via Ovid; EconLit via EBSCOhost; and CINAHL via EBSCOhost. The search strategies were developed by reviewing published HSUV review recommendations, the strategies of published HSUV protocols, and published guidance from other sources such as the Canadian Agency for Drugs and Technologies in Health (CADTH). Specifically, we will use two categories of search terms: disease-specific queries and health state utility elicitation methodology-specific queries. To strengthen this search strategy, we added utility-based instruments as an additional search query. To validate the search strategy, we reviewed the reference lists of published systematic reviews of health utility studies and collected 28 citations (the validation set) to reflect a range of years, methods, and cancer types[23-50] The search strategies developed for the current review were tested to confirm that they successfully captured the citations included in the validation set. The strategies were then adapted to the other databases. **Table 1** presents the search strategies in MEDLINE and EMBASE, which were searched separately.

Study eligibility

Inclusion criteria

A high-level set of inclusion criteria was selected for this review. To be eligible for inclusion, a publication must meet all the following:

- The publication presents the methods of a primary HSUV study, such that a study uses a direct elicitation method (e.g. TTO), an indirect method (e.g. EQ-5D), or a

combination of both to elicit preferences for health states from patients or non-patients (e.g. general public, family, caregivers, or clinicians);

- The study targets cancer, cancer treatments, and/or the cancer patient population, including non-patient respondents, in either a pediatric or adult setting; and
- The study reports HSUV estimates.

This review is limited to the context of patients who have been diagnosed with cancer, though studies have been published for related populations, such as unaffected high-risk individuals. No restrictions have been specified for publication date or language. In order to provide sufficient information to be used in CUAs, this review focuses only on peer-reviewed studies, excluding grey literature (e.g. unpublished studies, dissertations, conference abstracts). We anticipate that several reviews and economic evaluations, which reference HSUV literature, will be identified with the proposed search strategies. Reviewers will identify these records and the reference lists of these publications will be cross-referenced with the final list of included studies to assess the comprehensiveness of our review.

Data collection

Study selection

Prior to screening, duplicate publications will be identified and excluded. The titles and abstracts of all publications identified by the search strategies will be screened according to the eligibility criteria. Where unclear, reviewers will carry the record forward to the full-text screening phase. We have adopted this sensitive approach based on research demonstrating limitations in evaluating study eligibility at the title and

abstract level.[18] The full text publications of included abstracts will be retrieved and assessed for eligibility. Publications published in languages other than English will be reviewed by language-matched reviewers who having a working knowledge of the language of publication. All screening will be conducted independently and in duplicate. The flow of information process, which documents the number of records retrieved as well as the number of inclusions and exclusions at each screening phase, will be summarized in a PRISMA flow diagram.[51]

Based on preliminary searches, it is anticipated that a large volume of records will be retrieved through the literature search. Thus, this review necessitates the participation of multiple reviewers whose availability may change over time. As a means of promoting consistency across reviewers, we will establish a training set of 150 records purposefully chosen to represent a broad range of eligible and ineligible studies. Prior to beginning abstract screening, new reviewers will complete the training set and review their results, including reasons for exclusion, against the answer key. The complete set of publications to be screened will be divided into blocks of 1000 records. Reviewers will be assigned one block of records at a time and, upon completion, will be assigned a new, previously unscreened block. Once all blocks have been screened in single, the blocks will be re-assigned to satisfy the requirement of duplicate screening.

Data extraction and management

A complete list of data extraction items is presented in **Appendix A**. All data will be extracted independently and in duplicate using a similar blocking approach as described in the screening process. The reviewers who complete data extraction will

review and resolve discrepancies by discussion, with a third reviewer providing arbitration, as necessary. In the case of missing data, we will attempt to contact the corresponding authors for clarification. The data extraction form has been successfully piloted. All screening and data extraction will be maintained in Microsoft Excel workbooks which include extensive standardized vocabulary to promote consistency and ease of data extraction and reconciliation.

Data synthesis

A descriptive summary of the findings of this review will be presented, arranged by cancer type. Health state descriptions and corresponding HSUVs from each study will be presented along with a summary of the study methodologies and respondent characteristics. At present, this review is designed to gather and describe published HSUVs. Cognizant of the assumptions that must be made, particularly when HSUVs are derived through different methodologies, we will explore different quantitative evidence synthesis approaches that have been used to pool the HSUVs in the literature.[52] Based on this, we will make an informed decision on the synthesis approach for HUB.

Ethics and dissemination

No ethics approval will be sought for the purpose of this review as no primary data collection will take place. All information will be identified from published studies. The completed review will be disseminated in a series of publications in peer-reviewed journals, arranged by cancer type, detailing the systematic review methodology as well as a summary of the findings. We are also in the process of seeking funding support to develop an online portal to disseminate the HSUVs identified through this review.

Pilot review

Summary of screening and validation

To illustrate the systematic review process described here, we present the screening of a subset of records identified for thyroid cancer. According to data maintained by the World Health Organization (1970-2012) and the Cancer Incidence in Five Continents (1960-2007), the incidence of thyroid cancer has been increasing over the last several decades despite a falling mortality rate. These trends have been attributed to changes in the diagnosis, treatment, and exposure to risk factors.[53] However, if current trends persist, it is suggested that thyroid cancer may be the fourth most common cancer in the United States by the year 2030.[54] Despite this, our preliminary review suggested that there are relatively few published studies for health utilities for this indication.

From the complete set of records identified with our search strategy (N=52 551), we selected a subset that contained the phrase "thyroid" in the title or abstract (n=842). From these 842 abstracts, 31 were reviewed at the full-text screening level, and six fulfilled all eligibility criteria. Additionally, 21 economic evaluations were identified.[55-75] No systematic literature reviews related to health utilities for thyroid cancer were identified. The screening process is summarized in **Figure 1**.

From the 21 economic evaluations, 35 unique citations for health utility inputs were identified. To validate the systematic review process, these were cross-referenced with the list of included studies. Of these 35 citations, 12 were in a clinical area outside of thyroid cancer, ten lacked indexing or abstract keywords to indicate that health utility evidence was presented, four were published in a source not indexed in the included

medical literature databases (i.e. books, websites), and three referenced an economic evaluation where no original health utility study was conducted. The remaining six citations were also identified through our search strategy in thyroid cancer and were reviewed for inclusion through our systematic review. However, only two of these citations were considered eligible for inclusion [62,76]. While economic evaluations may contain bespoke health utility studies, they often lack the indexing or keywords to identify them as a health utility study. Overall, the validation steps suggest that the literature search strategy and screening process adequately identified all relevant publications in line with the pre-specified eligibility criteria.

Descriptive synthesis

Characteristics of the six eligible studies are presented in **Table 1**. Most studies reported on health utilities collected using a single technique except for one that employed multiple methodologies (EQ-5D; SF-6D; HUI-2; HUI-3).[77] Three studies used a cross-sectional design.[62,76,78] Respondents varied across studies, with three recruiting patients,[77-79] two recruiting clinicians,[62] and one recruiting members of the general public.[76] One health utility studied was conducted alongside a clinical trial.[79] However, participant characteristics were generally poorly described in the included publications. All health states were either derived for the purpose of the study or relied on patient's own health. Adverse events or toxicity was explicitly incorporated into health state descriptions only in one paper.[76] A summary of the published health utility estimates, arranged by respondent subgroups and scaling method, is presented in **Table 2**. The health state with the highest HSUV was obtained from clinicians using the TTO for

'Disease-free after thyroid lobectomy' (0.99),[62] while the lowest observed estimate was reported for 'Stable disease with grade 3 diarrhea' (0.42, SD 0.29, 95% CI: 0.36, 0.48) by the general public using the TTO.[76] In both studies, health state descriptions were provided by investigators.

Discussion

Published studies that measure HSUVs are a main source of health utilities used in CUAs. These evaluations are an integral component of reimbursement submissions prepared by drug manufacturers seeking listing on public formularies. However, the selection of HSUVs, where multiple studies are available, is often left to the discretion of analysts. Thus, the lack of a systematic approach to the identification and use of published health utilities may lead to a reimbursement policy that doesn't reflect the preferences of the public. The current review applies a systematic approach to the identification of published HSUVs and thus affords a level of confidence to knowledge users who rely on valid information to complete economic evaluation and HTA appraisals.

Where a health utility estimate does not exist in the literature for a given condition or health state, it is common to use estimates derived for a similar condition. This was the case in several of the economic evaluations identified through our pilot review. However, it is outside the scope of our review to suggest indications that may be interchangeable.

While the proposed review is extensive in scope, there are limitations. Estimates of HSUVs coming from grey literature sources, including conference abstracts and other unpublished media, are not eligible for inclusion. According to our past experience, information provided in conference abstracts or media reports often is not sufficient to be

used in CUAs. If the reporting in the grey literature changes in the future, we will revise our review to expand the search strategies and eligibility criteria accordingly. Despite this, the proposed review will culminate in a comprehensive summary of the evidence landscape for published HSUVs in oncology. Detailed study methodologies and respondent characteristics will be collected and summarized. Moreover, this review is the first component of the Health Utility Book (HUB) project.[80] The publication of this review protocol is in line with the HUB project team's goal to maintain transparency and accountability.

References

1. CADTH Common Drug Review. *Submission Guidelines for the CADTH Common Drug Review.* Canadian Agency for Drugs and Technology in Health,;2014.
2. Commonwealth of Australia. Procedure guidance for listing medicines on the Pharmaceutical Benefits Scheme. 2018.
3. National Institute for Health and Care Excellence. *Developing NICE guidelines: the manual.* 2014.
4. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England journal of medicine.* 1977;296(13):716-721.
5. Blinman P, King M, Norman R, Viney R, Stockler MR. Preferences for cancer treatments: an overview of methods and applications in oncology. *Annals of oncology : official journal of the European Society for Medical Oncology.* 2012;23(5):1104-1110.
6. Heymach J, Krilov L, Alberg A, et al. Clinical Cancer Advances 2018: Annual Report on Progress Against Cancer From the American Society of Clinical Oncology. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2018;36(10):1020-1044.
7. Vogler S, Vitry A, Babar ZU. Cancer drugs in 16 European countries, Australia, and New Zealand: a cross-country price comparison study. *The Lancet Oncology.* 2016;17(1):39-47.
8. Paying a high price for cancer drugs. *Lancet (London, England).* 2015;386(9992):404.
9. Brazier J. Valuing health States for use in cost-effectiveness analysis. *PharmacoEconomics.* 2008;26(9):769-779.
10. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14(1):102-109.

11. Tran AD, Fogarty G, Nowak AK, et al. A systematic review and meta-analysis of utility estimates in melanoma. *The British journal of dermatology.* 2018;178(2):384-393.
12. Jeong K, Cairns J. Systematic review of health state utility values for economic evaluation of colorectal cancer. *Health Economics Review.* 2016;6(1):36.
13. Paracha N, Thuresson PO, Moreno SG, MacGilchrist KS. Health state utility values in locally advanced and metastatic breast cancer by treatment line: a systematic review. *Expert review of pharmacoeconomics & outcomes research.* 2016;16(5):549-559.
14. Meregaglia M, Cairns J. A systematic literature review of health state utility values in head and neck cancer. *Health and quality of life outcomes.* 2017;15(1):174.
15. Forsythe A, Brandt PS, Dolph M, Patel S, Rabe APJ, Tremblay G. Systematic review of health state utility values for acute myeloid leukemia. *ClinicoEconomics and Outcomes Research: CEOR.* 2018;10:83-92.
16. Xie F, Zoratti M, Chan K, et al. 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). *Medical decision making : an international journal of the Society for Medical Decision Making.* 2019;39(4):370-378.
17. Papaioannou D, Brazier J, Paisley S. Systematic searching and selection of health state utility values from the literature. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2013;16(4):686-695.
18. Papaioannou D, Brazier J, Paisley S. NICE Decision Support Unit Technical Support Documents. In: *NICE DSU Technical Support Document 9: The Identification, Review and Synthesis of Health State Utility Values from the Literature.* London: National Institute for Health and Care Excellence (NICE) 2010.

19. Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. *PharmacoEconomics.* 2016;34(6):597-608.
20. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2012;15(5):639-649.
21. Kaltenthaler E, Tappenden P, Paisley S. Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2013;16(5):830-836.
22. Zechmeister-Koss I, Schnell-Inderst P, Zauner G. Appropriate evidence sources for populating decision analytic models within health technology assessment (HTA): a systematic review of HTA manuals and health economic guidelines. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2014;34(3):288-299.
23. Albertsen PC. Preferences of husbands and wives for outcomes of prostate cancer screening and treatment. *Journal of Urology.* 2005;173(2):565.
24. Albertsen PC, Nease Jr RF, Potosky AL. Assessment of patient preferences among men with prostate cancer. *Journal of Urology.* 1998;159(1):158-163.
25. Bennet CL, Chapman G, Elstein AS, et al. A comparison of perspectives on prostate cancer: Analysis of utility assessments of patients and physicians. *European Urology.* 1997;32(SUPPL. 3):86-88.
26. Best JH, Garrison LP, Hollingworth W, Ramsey SD, Veenstra DL. Preference values associated with stage III colon cancer and adjuvant chemotherapy. *Quality of Life Research.* 2010:1-10.
27. Beusterien KM, Davies J, Leach M, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: A cross-sectional utility study. *Health and quality of life outcomes.* 2010;8 (no pagination)(50).

28. Beusterien KM, Szabo SM, Kotapati S, et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. *British Journal of Cancer.* 2009;101(3):387-389.
29. Cappelli M, Surh L, Humphreys L, et al. Measuring women's preferences for breast cancer treatments and BRCA1/BRCA2 testing. *Quality of Life Research.* 2001;10(7):595-607.
30. Dominitz JA, Provenzale D. Patient preferences and quality of life associated with colorectal cancer screening. *American Journal of Gastroenterology.* 1997;92(12):2171-2178.
31. Grann VR, Jacobson JS, Sundararajan V, Albert SM, Troxel AB, Neugut AI. The quality of life associated with prophylactic treatments for women with BRCA1/2 mutations. *Cancer Journal from Scientific American.* 1999;5(5):283-292.
32. Havrilesky LJ, Broadwater G, Davis DM, et al. Determination of quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. *Gynecologic Oncology.* 2009;113(2):216-220.
33. Hayman JA, Fairclough DL, Harris JR, Weeks JC. Patient preferences concerning the trade-off between the risks and benefits of routine radiation therapy after conservative surgery for early- stage breast cancer. *Journal of Clinical Oncology.* 1997;15(3):1252-1260.
34. Hess LM, Malone DC, Reed PG, Skrepnek G, Weihs K. Preferences of patients and oncologists for advanced ovarian cancer treatment-related health States. *Health Outcomes Research in Medicine.* 2010;1(1):e51-e59.
35. Hornbrook MC, Wendel CS, Coons SJ, et al. Complications among colorectal cancer survivors: SF-6D preference-weighted quality of life scores. *Medical Care.* 2011;49(3):321-326.
36. Hudgens S, Briggs A, Tremblay G, Forsythe A, Lloyd A. Comparison of methods to estimate health state utilities in metastatic breast cancer (MBC). *Value in Health.* 2014;17 (7):A557.

37. Ko CY, Maggard M, Livingston EH. Evaluating health utility in patients with melanoma, breast cancer, colon cancer, and lung cancer: A nationwide, population-based assessment. *Journal of Surgical Research.* 2003;114(1):1-5.
38. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. *British Journal of Cancer.* 2006;95(6):683-690.
39. Lloyd A, Van Hanswijck De Jonge P, Doyle S, Cornes P. Health state utility scores for cancer-related anemia through societal and patient valuations. *Value in Health.* 2008;11(7):1178-1185.
40. Nafees B, Lloyd AJ, Dewilde S, Rajan N, Lorenzo M. Health state utilities in non-small cell lung cancer: An international study. *Asia Pacific Journal of Clinical Oncology.* 2016.
41. Nafees B, Patel C, Ray D, Gray E, Lau HJ, Lloyd AJ. An assessment of health-state utilities in metastatic breast cancer in the United Kingdom. *Value in Health.* 2016;19 (3):A157.
42. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health and quality of life outcomes.* 2008;6 (no pagination)(84).
43. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. *American Journal of Gastroenterology.* 1999;94(6):1650-1657.
44. Papatheofanis FJ. Utility evaluations for Markov states of lung cancer for PET-based disease management. *Quarterly Journal of Nuclear Medicine.* 2000;44(2):186-190.
45. Saigal CS, Gornbein J, Nease R, Litwin MS. Predictors of utilities for health states in early stage prostate cancer. *Journal of Urology.* 2001;166(3):942-946.
46. Saigal CS, Gornbein J, Reid K, Litwin MS. Stability of time trade-off utilities for health states associated with the treatment of prostate cancer. *Quality of Life Research.* 2002;11(5):405-414.
47. Shiroiwa T, Fukuda T, Tsutani K. Health utility scores of colorectal cancer based on societal preference in Japan. *Quality of Life Research.* 2009;18(8):1095-1103.

48. Smith DS, Krygiel J, Nease Jr RF, Sumner IW, Catalona WJ. Patient preferences for outcomes associated with surgical management of prostate cancer. *Journal of Urology.* 2002;167(5):2117-2122.
49. Stein D, Joulain F, Naoshy S, et al. Assessing health-state utility values in patients with metastatic colorectal cancer: a utility study in the United Kingdom and the Netherlands. *International Journal of Colorectal Disease.* 2014;29(10):1203-1210.
50. Volk RJ, Cantor SB, Cass AR, Spann SJ, Weller SC, Krahn MD. Preferences of husbands and wives for outcomes of prostate cancer screening and treatment. *Journal of General Internal Medicine.* 2004;19(4):339-348.
51. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Clinical research ed).* 2009;339:b2535.
52. Peasgood T, Brazier J. Is Meta-Analysis for Utility Values Appropriate Given the Potential Impact Different Elicitation Methods Have on Values? *PharmacoEconomics.* 2015;33(11):1101-1105.
53. La Vecchia C, Malvezzi M, Bosetti C, et al. Thyroid cancer mortality and incidence: a global overview. *International journal of cancer.* 2015;136(9):2187-2195.
54. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer research.* 2014;74(11):2913-2921.
55. Venkatesh SP, J. D.; Beninato, T.; Drake, F. T.; Kluijfhout, W. P.; Liu, C.; Gosnell, J. E.; Shen, W. T.; Clark, O. H.; Duh, Q. Y.; Suh, I. Cost-effectiveness of active surveillance versus hemithyroidectomy for micropapillary thyroid cancer. *Surgery.* 2017;161(1):116-126.

56. Al-Qurayshi ZK, E.; Randolph, G. W. Cost-effectiveness of intraoperative nerve monitoring in avoidance of bilateral recurrent laryngeal nerve injury in patients undergoing total thyroidectomy. *Br J Surg.* 2017;104(11):1523-1531.
57. Garcia AP, B. J.; Parks, N. A.; Liu, T. H. Routine prophylactic central neck dissection for low-risk papillary thyroid cancer is not cost-effective. *Clin Endocrinol (Oxf).* 2014;81(5):754-761.
58. Zanocco KH, M.; Elaraj, D.; Sturgeon, C. Cost effectiveness of intraoperative pathology examination during diagnostic hemithyroidectomy for unilateral follicular thyroid neoplasms. *J Am Coll Surg.* 2013;217(4):702-710.
59. Heller MZ, K.; Zydowicz, S.; Elaraj, D.; Nayar, R.; Sturgeon, C. Cost-effectiveness analysis of repeat fine-needle aspiration for thyroid biopsies read as atypia of undetermined significance. *Surgery.* 2012;152(3):423-430.
60. Wang TSC, K.; Mehta, P.; Roman, S. A.; Walker, H. D.; Sosa, J. A. To stimulate or withdraw? A cost-utility analysis of recombinant human thyrotropin versus thyroxine withdrawal for radioiodine ablation in patients with low-risk differentiated thyroid cancer in the United States. *J Clin Endocrinol Metab.* 2010;95(4):1672-1680.
61. Mernagh PC, S.; Dietlein, M.; Luster, M.; Mazzaferri, E.; Weston, A. R. Cost-effectiveness of using recombinant human TSH prior to radioiodine ablation for thyroid cancer, compared with treating patients in a hypothyroid state: the German perspective. *Eur.* 2006;155(3):405-414.
62. Esnaola NFC, S. B.; Sherman, S. I.; Lee, J. E.; Evans, D. B. Optimal treatment strategy in patients with papillary thyroid cancer: a decision analysis. *Surgery.* 2001;130(6):921-930.
63. Zanocco KK, D. J.; Wu, J. X.; Fingeret, A.; Heller, K. S.; Lee, J. A.; Yeh, M. W.; Sosa, J. A.; Sturgeon, C. Cost Effectiveness of Routine Laryngoscopy in the Surgical Treatment of Differentiated Thyroid Cancer. *Annals of Surgical Oncology.* 2018:1-8.

64. Lang BHHW, C. K. H. Lobectomy is a more Cost-Effective Option than Total Thyroidectomy for 1 to 4 cm Papillary Thyroid Carcinoma that do not Possess Clinically Recognizable High-Risk Features. *Annals of Surgical Oncology.* 2016;23(11):3641-3652.
65. Balentine CJV, D. J.; Schneider, D. F. Cost-effectiveness of lobectomy versus genetic testing (Afirma) for indeterminate thyroid nodules: Considering the costs of surveillance. *Surgery (United States).* 2018;163(1):88-96.
66. Wilson LH, W.; Chen, L.; Ting, J.; Cao, V. Cost Effectiveness of Lenvatinib, Sorafenib and Placebo in Treatment of Radioiodine-Refractory Differentiated Thyroid Cancer. *Thyroid.* 2017;27(8):1043-1052.
67. Huang WC, L.; Cao, V.; Sung, H.; Yokokura, M.; Ting, J.; Wilson, L. Cost effectiveness of lenvatinib, sorafenib, and placebo in treatment of radioiodine-refractory differentiated thyroid cancer. *Value in Health.* 2016;19 (3):A204.
68. Tremblay GP, C.; Copher, R.; Forsythe, A.; Majethia, U. Cost-effectiveness analysis of lenvatinib as a treatment for radioactive iodine refractory differentiated thyroid cancer in The United States. *Value in Health.* 2016;19 (3):A151.
69. Stern SH, O.; Horowitz, E.; Leshno, M.; Feinmesser, R. Is There Justification for Total Thyroidectomy in Low-Risk Papillary Thyroid Carcinoma? A Decision-Analysis Model. *World journal of surgery.* 2015;39(11):2707-2717.
70. Erdal ES, H.; Turkmen, C.; Aral, F.; Yildiz, O.; Okutur, K.; Parali, E.; Deger, C.; Tunalioglu, A.; Sar, C.; Asan, S.; Sumer, F.; Ozel, O. Cost-effectiveness of sorafenib for treatment of radioactive iodine (RAI)-refractory locally advanced/metastatic differentiated thyroid cancer (DTC) in Turkey. *Value in Health.* 2015;18 (3):A203-A204.
71. Vriens DA, E. M. M.; Netea-Maier, R. T.; Smit, J. W. A.; De Wilt, J. H. W.; Oyen, W. J. G.; De Geus-Oei, L. Cost-effectiveness of FDG-PET/CT for cytologically indeterminate thyroid nodule. *European Journal of Nuclear Medicine and Molecular Imaging.* 2013;2):S239-S240.

72. Mailhot Vega RK, J.; Bussiere, M.; Hattangadi, J.; Hollander, A.; Michalski, J.; Tarbell, N.; Yock, T.; MacDonald, S. Cost-effectiveness of proton therapy compared to photon therapy in the management of pediatric medulloblastoma. *International Journal of Radiation Oncology Biology Physics.* 2013;1):S5.
73. Borget IB, J.; Catargi, B.; Deandreis, D.; Zerdoud, S.; Bridji, B.; Bardet, S.; Leenhardt, L.; Bastie, D.; Schvartz, C.; Vera, P.; Morel, O.; Benisvy, D.; Bournaud, C.; Bonichon, F.; Dejax, C.; Toubert, M.; Ricard, M.; Leboulleux, S.; Benhamou, E.; Schlumberger, M. Cost effectiveness of strategies of radioiodine ablation in thyroid carcinoma patients: Results of the randomized phase III estimabl study. *Thyroid.* 2012;1):A63.
74. Zanocco KP-z, L.; Dalal, S.; Elaraj, D.; Nayar, R.; Sturgeon, C. On-site adequacy evaluation is not cost effective for experienced operators performing initial ultrasound-guided fine-needle aspiration of thyroid nodules. *Thyroid.* 2012;1):A44.
75. Zanocco KS, C. The cost-effectiveness of recombinant human thyroid stimulating hormone administration prior to remnant ablation for treatment of differentiated thyroid cancer. *Journal of Surgical Research.* 2012;172 (2):239.
76. Fordham BAK, C.; de Freitas, H. M.; Lloyd, A. J.; Johnston, K.; Pelletier, C. L.; Tremblay, G.; Forsythe, A.; McIver, B.; Cohen, E. E. W. Health state utility valuation in radioactive iodine-refractory differentiated thyroid cancer. *Patient Preference and Adherence.* 2015;9:1561-1572.
77. Lubitz CCDG, L.; Fingeret, A. L.; Economopoulos, K. P.; Termezawi, D.; Hassan, M.; Parangi, S.; Stephen, A. E.; Halpern, E. F.; Donelan, K.; Swan, J. S. Measurement and Variation in Estimation of Quality of Life Effects of Patients Undergoing Treatment for Papillary Thyroid Carcinoma. *Thyroid.* 2017;27(2):197-206.
78. Kent EEA, A.; Mitchell, S. A.; Clauser, S. B.; Smith, A. W.; Hays, R. D. Health-related quality of life in older adult survivors of selected cancers: Data from the SEER-MHOS linkage. *Cancer.* 2015;121(5):758-765.

79. Borget IB, J.; Catargi, B.; Deandreis, D.; Zerdoud, S.; Rusu, D.; Bardet, S.; Leenhardt, L.; Bastie, D.; Schvartz, C.; Vera, P.; Morel, O.; Benisvy, D.; Bournaud, C.; Bonichon, F.; Kelly, A.; Toubert, M. E.; Leboulleux, S.; Journeau, F.; Benhamou, E.; Schlumberger, M. Quality of life and cost-effectiveness assessment of radioiodine ablation strategies in patients with thyroid cancer: Results from the randomized phase III ESTIMABL trial. *Journal of Clinical Oncology.* 2015;33(26):2885-2892.

80. Xie F. 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). 2018.

Tables and Figures

Table 1: Search strategy for MEDLINE and EMBASE (via Ovid)

Table 2: Characteristics of included studies of health utility in thyroid cancer

Table 3: Health states in thyroid cancer

Figure 1: PRISMA flow of information diagram

Table 1: Search strategy for MEDLINE and EMBASE (via Ovid)

Search	Query
Cancer-related search terms	
1	cancer*.mp.
2	exp Neoplasms/
3	neoplasm*.mp.
4	exp Carcinoma/
5	carcinoma*.mp.
6	exp Sarcoma/
7	sarcoma*.mp.
8	exp Lymphoma/
9	lymphoma*.mp.
10	exp Leukemia/
11	leukemia*.mp.
12	myeloma.mp.
13	tumor*.mp.
14	tumour*.mp.
15	Or/1-14
Health utility elicitations methods-related search terms	
16	(health adj3 (utilit* or status)).mp.
17	(utilit* adj3 (valu* or measur* or health or life or estimat* or elicit* or disease or score* or weight)).mp.
18	(preference* adj3 (valu* or measur* or health or life or estimat* or elicit* or disease or score* or instrument* or scale* or quest*)).mp.
19	disutilit*.mp.
20	standard gamble*.mp.
21	(time trade off or time tradeoff or time trade-off).mp.
22	tto.ti,ab,kw.
23	visual analog* scale*.mp.
24	VAS.mp.
25	discrete choice experiment*.mp.
26	Rating scale.mp. AND (health adj3 (utilit* or status)).mp.
27	(Personal trade-off or PTO).mp AND (health adj3 (utilit* or status)).mp.
28	(multiattribute health status* or multi-attribute health status* or multiattribute utility* or multi-attribute utility*).mp.
Utility-based instrument search terms	
29	(hui or hui1 or hui2 or hui3).ti,ab,kw.
30	health utility index.mp.
31	(eq or euroqol or euro qol or eq5d or eq 5d or euroqual or euro qual or European Quality of Life 5-dimension or EQ-5D or EQ5D or EQ 5D).mp.
32	(sf6d or sf 6d or sf-6d or short form 6d or shortform 6d).mp.

Search	Query
33	(15-D or 15D).mp AND (health utility or health utilities or utility or utilities).mp
34	(AQoL or AQL or Assessment of Quality of Life).mp
35	(Patient ORiented Prostate Utility Scale or PORPUS).mp
36	(PROMIS or Patient-Reported Outcomes Measurement Information System).mp AND (health utility or health utilities or utility or utilities).mp
37	Or/16-36
38	**15 and 37**

Table 2: Characteristics of included studies of health utility in thyroid cancer

Study	Scaling method	Respondents	Mode of administration	Source of health state descriptions	Country
Lubitz et al., 2017	EQ-5D SF-6D HUI-2 HUI-3	Patients	Trained interviewer or mail; Subsequent surveys conducted online	Own health (implied)	USA
Esnaola et al., 2001	TTO	Clinicians	Unclear	Investigator-proposed	USA
Kent et al., 2015	SF-6D	Patients	Mail or telephone	Own health (implied)	USA
Fordham et al., 2015	TTO	General public	Face-to-face interviews	Vignettes designed through a qualitative study with patients and clinicians	United Kingdom
Borget et al., 2015	EQ-5D	Patients	Unclear	Own health (implied)	France

Table 3: Health states in thyroid cancer

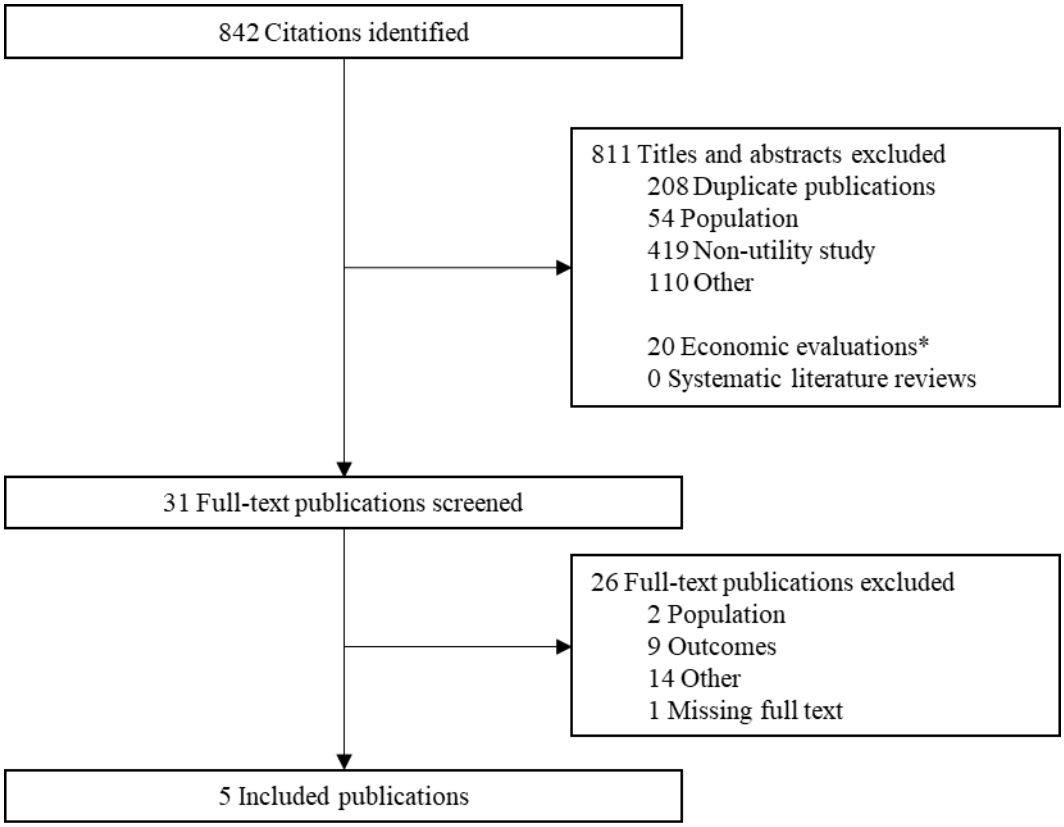
Author	Health state	Respondents	Scaling method	Analysis N	Mean (SD) Median (Q1, Q3)
Lubitz et al., 2017	Pre-operation	All participants	EQ-5D	117	0.895 (0.103) 0.876 (0.82, 1)
			SF-6D	117	0.773 (0.125) 0.793 (0.66, 0.86)
			HUI-2	117	0.875 (0.133) 0.917 (0.83, 0.95)
			HUI-3	117	0.859 (0.185) 0.919 (0.79, 0.97)
	Post-operation	All participants	EQ-5D	117	0.882 (0.114, 95% CI: 0.665, 1)
			SF-6D	117	0.748 (0.117, 95% CI: 0.548, 0.919)
			HUI-2	117	0.873 (0.120, 95% CI: 0.647, 1)
			HUI-3	117	0.843 (0.167, 95% CI: 0.518, 1)
	26 weeks post-operation	All participants	EQ-5D	117	0.911 (0.107, 95% CI: 0.752, 1)
			SF-6D	117	0.798 (0.122, 95% CI: 0.609, 0.922)
			HUI-2	117	0.879 (0.111, 95% CI: 0.705, 1)
			HUI-3	117	0.863 (0.136, 95% CI: 0.596, 1)
Esnaola et al., 2001	Disease-free after thyroid lobectomy	All participants	TTO	15	0.99
	Disease-free after total thyroidectomy/radioiodine therapy			15	0.95
	Disease-free after thyroid surgery/permanent complication			15	0.88
	Disease-free after surgery for cervical recurrence			15	0.95

Author	Health state	Respondents	Scaling method	Analysis N	Mean (SD) Median (Q1, Q3)
	Systemic recurrence			15	0.60
Kent et al., 2015	Thyroid cancer	All participants	SF-6D	386	0.70 (95% CI: 0.69, 0.71)
Fordham et al., 2015	Stable/no response	All participants	TTO	100	0.80 (0.19, 95% CI: 0.77, 0.84)
	Response to therapy				0.86 (0.15, 95% CI: 0.83, 0.89)
	Progressive disease				0.50 (0.28, 95% CI: 0.45, 0.56)
	Stable + Grade 3 Diarrhea				0.42 (0.29, 95% CI: 0.36, 0.48)
	Stable + Grade 3 fatigue				0.72 (0.24, 95% CI: 0.67, 0.77)
	Stable + Grade 3 Hand and foot syndrome				0.52 (0.30, 95% CI: 0.46, 0.58)
	Stable + Grades 1 or 2 alopecia				0.75 (0.21, 95% CI: 0.71, 0.79)
Borget et al., 2015	At treatment assignment	THW	EQ-5D	336	0.87
	Immediately before radioiodine administration	THW		336	0.84
	2 weeks	THW		336	0.82
	4 weeks	THW		336	0.85
	6 weeks	THW		336	0.87
	3 months	THW		336	0.88
	8 months	THW		336	0.90
	At treatment assignment	rhTSH		348	0.84
	Immediately before radioiodine administration	rhTSH		348	0.85
	2 weeks	rhTSH		348	0.86
	4 weeks	rhTSH		348	0.86
	6 weeks	rhTSH		348	0.87
	3 months	rhTSH		348	0.86

Author	Health state	Respondents	Scaling method	Analysis N	Mean (SD) Median (Q1, Q3)
	8 months	rhTSH		348	0.88
	At radioidine administration	THW		336	0.833 (0.192)
		rhTSH		348	0.849 (0.173)
		3.7 GBq radioidine activity		337	0.836 (0.184)
		1.1 GBq radioidine activity		347	0.846 (0.182)

THW: Patients managed with thyroid hormone withdrawal; rhTSH – Patients managed with recombinant human thyroid-stimulation hormone

Figure 1: PRISMA flow of information diagram



The 842 citations screened in this pilot were identified as containing the term “thyroid” from the set of 52,551 records identified through the primary review’s search strategy.

*One record was flagged as both an include at the title/abstract screening level and as an economic evaluation

Appendix A

Complete list of data extraction items, arranged by general data category.

Study Overview

- Study identifiers
    - Author
    - Year
    - Title
    - Journal
    - Conference abstract only?
    - Technique

- Respondents
    - Respondents
    - Country(s)

- Disease context
    - Cancer type
    - Number staging
    - TMN staging
    - Genotype variation
    - Line(s) of treatment
    - Current treatment(s)
    - Previous treatment(s)

- Study design
    - Is the HSUV study part of a clinical trial?
        - Trial name and identifier(s)
        - Trial phase
    - Cross-sectional or longitudinal
    - Total study duration (weeks)
    - Number of data collection points
    - Subgroups with data

- Indifference procedures
    - Matching or indifference procedure used?
        - If yes, procedure used

- Health states
    - Source of health state description (New, Existing, Mixed)
    - If health states sourced from the literature, provide citations
    - If new health states, describe method for deriving
    - Reference treatment(s) for health states

    - Anchors defined?
    - Defined lower anchor
    - Value of lower bound
    - Defined upper anchor
    - Value of upper bound
    - Order of health state presentation defined?
    - Visual aid used?
    - Describe any visual aids used

- Other assumptions

Methods & Results

- Study identifiers
    - Record ID
    - Author
    - Year
    - Technique
    - Respondents
    - Subgroup
    - Data collection time point (weeks)

- Administration
    - Mode of administration
    - Source of administrator
    - Training
    - Software package used?
    - If software used, name
    - If software used, was an interviewer present?
    - Responses recorded electronically or pen/paper
    - Responses record by (patient/investigator/other/unclear)
    - Training to respondent (any approach)

- Number of respondents
    - Number of respondents recruited
    - Number of respondents completing task(s)
    - Number of respondents included in analysis
    - Sample size calculations, if provided
        - Calculated N
        - Justification for N

- Respondent characteristics
    - Age
    - Sex
    - Race/ethnicities

    - Education
    - Income
    - Clinical characteristics, if respondents have condition of interest
        - Cancer type
        - Staging
        - Genotype variation
        - Line of treatment
        - Previous treatment
        - Current treatment
    - Language of study

- HSUV estimates (1 per health state)
    - Name of health state
    - Verbatim description
    - Duration of health state
    - N
    - Mean, Median, SD, Min, Max, Lower 95% CI, Upper 95% CI
    - Toxicity component to health state?
    - Name of toxic component
    - Duration of toxicity

CHAPTER 2-2 – A SYSTEMATIC LITERATURE REVIEW OF HEALTH STATE UTILITY VALUES IN BLADDER CANCER

Status: At the time of thesis submission, this manuscript has not been submitted for publication.

A systematic literature review of health state utility values in bladder cancer

Abstract

Background: Health utilities are integral inputs to the cost utility analyses which many decision makers rely on worldwide. The Health Utility Book (HUB) was designed to provide tools and means by which investigators may more readily identify, appraise, and apply the health utility literature to decision problems. The generation of a registry of health utility estimates in cancer is a pillar of the HUB project.

Methods: As part of the HUB Cancer project, a systematic literature review was conducted to identify studies collecting health utility outcomes in bladder cancer. Searches were executed in MEDLINE, EMBASE, CINAHL, and EconLit in March 2018. Titles and abstracts, as well as the full text publications of included records, were screened independently and in duplicate by at least two reviewers. Data extraction from included publications was conducted independently and in duplicate by two reviewers using piloted extraction templates. Under the HUB Cancer review protocol, screening was conducted for all cancer types simultaneously, with data extraction conducted through a dedicated process for each tumour type. Findings were summarized through a narrative synthesis.

Results: From the 64,231 records identified through the medical literature databases, 607 publications were included in the HUB Cancer review. With the addition of a single study identified through hand searches, six studies were included in the summary of evidence for bladder cancer. All studies reported health utility estimates collected from questionnaires completed by patients based on their own health status and experience. Five studies reported health utility estimates related to bladder cancer as part of a more

broadly scoped research objective, with one study describing outcomes based on treatment experience and age. All studies reported at least one health utility value for a broadly defined bladder cancer health state, with values varying from 0.683 with the SF-6D to 0.80 with the EQ-5D-3L.

Discussion: Limited evidence was identified for health utility in bladder cancer, with most studies only reporting estimates for a broadly defined health state. The single study which evaluated utility estimates by treatment and age suggested that important differences may persist across these stratifications. Over the time period in which studies were conducted, estimates for bladder cancer remained stable. Further research, either through integration into clinical trials or the conduct of bespoke investigations, is warranted to more comprehensively inform the health utility parameters in future cost utility analyses for bladder cancer.

Background

Cancer is commonly cited as a leading contributor to the collective public health and economic burdens faced by societies globally. Bladder cancer ranks among the top ten most commonly diagnosed types of cancers, with some 573,278 new cases diagnosed in the year 2020 according to the World Health Organization's International Agency for Research on Cancer.[1] Incidence varies by geography, with notably higher cases in Europe and North America,[2] though mortality estimates remain relatively consistent worldwide.[1] However, while GLOBOCAN data suggests an overall five-year survival of 77% in the United States, this estimate drops to 5% for patients with metastatic disease.[3] Importantly, patients with bladder cancer tend to report significant declines in health-related quality of life.[4-7] Citing factors such as high recurrence rates, intensive surveillance programs, and high-cost treatments, the management of patients with bladder cancer is among the highest of all cancer types.[8-10]

As has been described elsewhere, the quality-adjusted life year (QALY) is an outcome designed to reflect both mortality and health-related quality of life in a single index.[11] Health utility values are preference weights used to quantitatively describe health states and, by convention, have an upper bound of 1, used to denote full or perfect health, and a lower bound of 0, used for dead. Though not commonly used in practice, some scales theoretically allow health utility values to be less than 0 for health states considered to be worse than dead. The QALY is found by summing the product of time one spends in a health state and the health utility weight corresponding to that health state.

Globally, decision makers across several jurisdictions rely on economic evaluations to support reimbursement decisions. The QALY's portability across disease and intervention contexts has made it a valuable tool in the systematic evaluation of health technologies as this metric is readily comparable across decision problems. Economic evaluations which use the QALY to reflect estimates of effect for the interventions being compared are labelled 'cost utility analyses'. As integral parameters to these evaluations, a comprehensive approach to the identification, selection, and appraisal of health utility estimates is essential to ensure that the best available evidence is being considered. As a part of the Health Utility Book (HUB) project, an initiative designed to support investigators in identifying, appraising, and applying health utility evidence in decision making, the objective of this systematic literature review was to identify and describe the health utility evidence for bladder cancer.

Methods

The methods of this systematic literature review have been described previously.[12] Briefly, search strategies were adapted from published guidance and previously published literature reviews to include controlled vocabulary as well as keywords and text terms for disease-specific and methodology-specific health utility queries. Searches were executed in MEDLINE (via Ovid), EMBASE (via Ovid), EconLit (via EBSCOhost), and CINAHL (via EBSCOhost). To be eligible for inclusion, a publication must have presented the results of a primary health utility study for patients with or respondents completing preference-eliciting exercises on the topic of bladder cancer or from a secondary analysis of a previous study or database. Eligibility was restricted to full-text publications.

The titles and abstracts of identified records were screened independently and in duplicate by two reviewers. If either reviewer included the abstract, the full text publication was retrieved and evaluated against the complete eligibility criteria. Discrepancies between reviewers were reviewed and resolved by the project leads (MZ, FX). Systematic literature reviews of health utility studies and cost utility analyses were flagged by reviewers to be used in cross-referencing and validation steps. Data elements, including study methodologies, respondent characteristics, and health utility estimates, were extracted independently and in duplicate by two reviewers using piloted data extraction templates.

Screening of abstracts and full-text publications was completed for all cancer types simultaneously. Data extraction for publications reporting health utility estimates for bladder cancer was completed through a separate, dedicated process.

Given the heterogeneity across health utility studies, no statistical synthesis was planned to summarize the evidence identified through this review.

Results

Systematic literature review

Searches were executed on March 19, 2018. From the initial 64,231 records identified from the systematic searches of the bibliographic databases, 62,014 publications were excluded at the abstract screening phase with a further 1,663 publications excluded during full-text screening. The 607 included publications were sorted into 19 cancer-type categories, with five records tagged as reporting health utility estimates for bladder cancer.[13-17] Across the screening phases, 1,240 records were flagged

as either a systematic review of health utility studies or a cost utility analysis. Eight of these records were reviewed for relevant bladder cancer citations. To increase the sensitivity of the title and abstract screening stage, records may have been flagged both as an include and as a review of health utility studies or a cost utility analysis. One additional study was included through the cross-referencing exercise.[18] Overall, six studies were included for the summary of health utility estimates in bladder cancer (**Table 1; Table 2**).[13-18] The review process is illustrated in **Figure 1**.

Studies reporting health utilities in bladder cancer

Sullivan et al., 2011

Citing the success and the acceptance of the United States-based catalogue of EQ-5D scores[19] and a lack of a catalogue of preference-based health-related quality of life scores for the United Kingdom, Sullivan and colleagues applied United Kingdom tariffs to the EQ-5D-3L responses from the (MEPS) to derive health utility estimates for International Classification of Diseases, Ninth Revision (ICD-9) and Clinical Classification Categories (CCC) codes. The MEPS was a nationally representative survey of United States adults which collected, among other data, medical condition diagnoses and responses to health status surveys.[20] Investigators used data collected from the year 2000 to 2003 to include 79,522 unique individuals. Based on ICD-9 code 188 ("Malignant neoplasm of bladder"), the mean EQ-5D index value was 0.71 (95% CI: 0.65, 0.78, n = 63). Similarly, based on CCC code 032 ("Cancer of Bladder"), the mean EQ-5D index value was 0.71 (95% CI: 0.65, 0.77, n = 72). Under the ICD-9 and CCC codes, respectively, respondents with bladder cancer had an average of 5.4 and 5.5 total

chronic conditions. When framed as a disutility and controlled for age, comorbidity, gender, race, ethnicity, income, and education, the impact of bladder cancer was reported as -0.06 (95% CI: -0.11, -0.01) based on the ICD-9 classification and -0.04 (95% CI: -0.09, 0.01) based on the CCC code.

Hays et al., 2014

The impact of ten cancers and 13 other chronic medical conditions were described in terms of health-related quality of life based on data collected from 126,366 respondents from 1998 to 2002. The study sample was derived through linkage between the surveillance, epidemiology, and end results (SEER) cancer registry and the Medicare Health Outcomes Survey (MHOS) in the United States. Across all patients whose data was used in this study, the mean SF-6D score was 0.73 (SD 0.14), though patients with cancer comprised a minority of this sample (n = 22,740, 18.0%). Health utility estimates for patients with bladder cancer (n = 1,299) were presented based on two adjusted analyses. In the first analysis, based on recycled predictions with other independent variables fixed at their means, the mean health utility was 0.722. In a second analysis, where other conditions were fixed at zero and other independent variables were fixed at their means, the mean health utility was 0.793.

Kent et al., 2015

The SEER-MHOS data linkage was also used by Kent and colleagues (2015) to describe health-related quality of life for a population-based study of patients with and without cancer in the United States. For patients with cancer, analyses were based on the first survey after diagnosis. Bladder cancer was the most common tumour site represented

in the sample (n = 3,195) where patients were diagnosed at a mean age of 70.1 (SD 8.9) years. On average, patients with bladder cancer completed the survey 86.2 (SD 76.7) month after diagnosis. The average health utility for patients with bladder cancer was estimated as 0.70 (95% CI: 0.69, 0.70) according to the SF-6D/VR-6D based on a sample of 2,035 patients.

Hever et al., 2015

To validate the Hungarian version of the Bladder Cancer Index, Hevér and colleagues (2014) enrolled 151 patients with bladder cancer from three hospital-based urology centers in Hungary between May 2012 and September 2013. Patients also completed the EQ-5D and SF-36. SF-6D utility scores were derived from the SF-36 using ordinal, standard gamble, Bayesian, and parametric approaches. Citing a lack of local, Hungary-based value sets, investigators used United Kingdom tariffs to calculate utility scores. General health utility estimates for patients with bladder cancer varied from 0.683 (SD 0.136, n = 125) according to the Bayesian posterior mean to 0.784 (SD 0.242, n = 148) based on the EQ-5D. Additional treatment-specific estimates were available, including ileal conduit cystectomy, neobladder cystectomy, transurethral resection, and transurethral resection intravesical therapy, as well as overall age-stratified health utility scores measured by the EQ-5D.

Pickard et al., 2016

Health-related quality of life outcomes, including health utility estimates by the EQ-5D, were estimated in a retrospective analysis of a United States-based clinical trial which enrolled patients with 11 different types of advanced cancer. Patients had received

at least two cycles of chemotherapy or, if the dosing schedule was not cyclical, at least one month of treatment. A target sample size of 50 patients per cancer type was defined, though the target enrollment for bladder cancer was not achieved (n = 31). Using a United States preferencing-based algorithm for the EQ-5D, the mean health utility for patients with bladder cancer was 0.81 (SD 0.13). For reference, the mean EQ-5D health utility value for all patients was 0.78 (SD 0.15).

Naik et al., 2017

To establish a set of health utility scores for cancer sites based on Canadian preference weights, Naik and colleagues (2017) enrolled 1,759 cancer survivors in a cross-sectional study between May 2012 and December 2014. In addition to providing access to medical records and completing a demographic survey, participants also independently completed EQ-5D-3L questionnaires. Health utility scores were calculated based on three preference weight algorithms, including for Canada, the United Kingdom, and the United States. Patients with a history of bladder cancer represented a minority of respondents (n = 13/1759, 1%). Utility estimates varied based on the algorithm applied, with a mean 0.74 (SEM 0.06) based on the United Kingdom set, 0.77 (SEM 0.04) based on the Canadian set, and 0.80 (SEM 0.04) based on the United States preference algorithm.

Discussion

Across the six included studies, health utility estimates were derived from a collective 3,526 patients. Point estimates for the general definition of bladder cancer were generally consistent across studies and varied from approximately 0.68 to 0.81. Most

studies (5/6, 83%) reported estimates as part of more broadly scoped research objectives, designed to enroll and describe health utility estimates across multiple primary tumour sites. In such studies, bladder cancer-specific characteristics such as disease stage, age, and the distribution of respondent sex or gender, were unavailable. In all studies, it was implied that respondents were patients with bladder cancer completing questionnaires based on their own health. Health states were broadly defined, with only the investigation by Hever and colleagues providing additional stratifications by respondent age and treatment.

Every included study reported a generalized estimate for patients with bladder cancer (**Figure 2**), which varied from 0.683 in the investigation by Hever and colleagues which reported based on the SF-6D to 0.80 in the study by Naik and colleagues using the EQ-5D-3L. Irrespective of the measure used, these data illustrate the negative impact a bladder cancer diagnosis has on patient quality of life. For instance, all point estimates fell below population norms for the United Kingdom (Mean 0.828, n = 79,522) and the United States (Mean 0.867, n = 38,678) based on data collected during the MEPS.[18,19]

The EQ-5D and the SF-6D were the only two measures used in the included studies. These multi-attribute utility instruments are commonly employed in clinical research as they are readily accessible to investigators and are simple for respondents to complete with minimal direction or oversight. Direct comparisons of the health utility values estimated with different instrument could only be inferred from the results of the investigation by Hever and colleagues. The SF-6D estimates for the more broadly defined bladder cancer health state, which were based on ordinal, standard gamble, Bayesian and

parametric analytical approaches, were consistently lower than the estimate generated from the EQ-5D. Despite the likely heterogeneity across studies, this trend was generally observed across the point estimates from each investigation.

While several health utility estimates were found for a broadly scoped definition of bladder cancer, the available data is largely non-specific to varying disease characteristics and interventions. Indeed, only a single study described health utility estimates based on patient age or treatment and important variations were observed across these stratifications. Here, Hever and colleagues posited the impact on health-related quality of life not only of the index condition, but of the therapies a patient may undergo. While the sample size on which these comparisons may be drawn is limited, there is evidence to suggest some difference in health utility estimates for patients who undergo different treatments. To more readily inform economic evaluations in this clinical context, more research on the impact of treatments as well as patient and disease characteristics is warranted. This research should be extended to evaluating the impact of bladder cancer-specific adverse events. Moreover, while it is common practice to infer transferability of health utility inputs across clinical contexts by drawing on similar patient experiences and the impact of disease characteristics of patient's health-related quality of life, such inferences necessitate assumptions and justification. Such a decision was explained in a cost-utility analysis by Kulkarni and colleagues (2009) who cited a lack of health utility data in bladder cancer to inform model parameters.[21] The need for leaps of faith in these situations is among the motivations for other components of HUB project, of which this review is a central pillar.[22]

This was an extensive systematic literature review where several medical literature databases were queried and the retrieved records were evaluated according to systematic review good practices. Extensive validation steps, including hand searches of the bibliographies of included publications as well as relevant reviews and economic evaluations, were taken to evaluate the completeness of this set of evidence. Although conference abstracts were not eligible for inclusion as they were not considered to contain sufficient information to fully evaluate their methodologies and results, it is anticipated that such publications would later appear in a full-text publication. A review of submissions to health technology assessment bodies may yield further evidence.

Conclusion

Mean estimates of health utility for bladder cancer remained stable over the time period in which the six studies included in this systematic review were conducted. Overall, however, there is limited evidence in this clinical space and further research may be needed to more comprehensively inform the health utility inputs in cost utility analyses. As the first summary to come from the HUB Cancer literature review, this study establishes some of the concepts and patterns that may be presented and discussed. Given the limited number of studies identified here, few inferences and conclusions may be made. As research continues to evolve, not only in bladder cancer but through the more rigorous incorporation of health-related quality of life, including health utility, outcomes in clinical trials, tendencies and patterns may emerge across stratifications, including disease stages, treatments, and geographies.

References

1. International Agency for Research on Cancer. Cancer Today. World Health Organization,. https://gco.iarc.fr/today/home. Published 2021. Accessed 13 June 2021.
2. Bray F, Ferlay J, Laversanne M, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. *International journal of cancer.* 2015;137(9):2060-2071.
3. Saginala K, Barsouk A, Aluru JS, Rawla P, Padala SA, Barsouk A. Epidemiology of Bladder Cancer. *Med Sci (Basel).* 2020;8(1):15.
4. Cox E, Saramago P, Kelly J, et al. Effects of Bladder Cancer on UK Healthcare Costs and Patient Health-Related Quality of Life: Evidence From the BOXIT Trial. *Clinical Genitourinary Cancer.* 2020;18(4):e418-e442.
5. Catto JWF, Downing A, Mason S, et al. Quality of Life After Bladder Cancer: A Cross-sectional Survey of Patient-reported Outcomes. *Eur Urol.* 2021;79(5):621-632.
6. Smith AB, Jaeger B, Pinheiro LC, et al. Impact of bladder cancer on health-related quality of life. *BJU Int.* 2018;121(4):549-557.
7. Mason SJ, Downing A, Wright P, et al. Health-related quality of life after treatment for bladder cancer in England. *British Journal of Cancer.* 2018;118(11):1518-1528.
8. Leal J, Luengo-Fernandez R, Sullivan R, Witjes JA. Economic Burden of Bladder Cancer Across the European Union. *Eur Urol.* 2016;69(3):438-447.
9. Mossanen M, Gore JL. The burden of bladder cancer care: direct and indirect costs. *Curr Opin Urol.* 2014;24(5):487-491.
10. Svatek RS, Hollenbeck BK, Holmäng S, et al. The economics of bladder cancer: costs and considerations of caring for this disease. *Eur Urol.* 2014;66(2):253-262.
11. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *The New England journal of medicine.* 1977;296(13):716-721.

12. Zoratti MJ, Zhou T, Chan K, et al. Health Utility Book (HUB)–Cancer: Protocol for a Systematic Literature Review of Health State Utility Values in Cancer. *MDM Policy & Practice.* 2019;4(2):2381468319852594.
13. Hays RDR, B. B.; Smith, A. W.; Clauser, S. B. Associations of cancer and other chronic medical conditions with SF-6D preference-based scores in Medicare beneficiaries. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2014;23(2):385-391.
14. Hever NVP, M.; Ballo, A.; Gulacsi, L.; Baji, P.; Brodszky, V.; Damasdi, M.; Bognar, Z.; Toth, G.; Buzogany, I.; Szanto, A. Health related quality of life in patients with bladder cancer: a cross-sectional survey and validation study of the Hungarian version of the Bladder Cancer Index. *Pathol Oncol Res.* 2015;21(3):619-627.
15. Kent EEA, A.; Mitchell, S. A.; Clauser, S. B.; Smith, A. W.; Hays, R. D. Health-related quality of life in older adult survivors of selected cancers: Data from the SEER-MHOS linkage. *Cancer.* 2015;121(5):758-765.
16. Naik HH, D.; Su, S.; Qiu, X.; Brown, M. C.; Vennettilli, A.; Irwin, M.; Pat, V.; Solomon, H.; Wang, T.; Hon, H.; Eng, L.; Mahler, M.; Thai, H.; Ho, V.; Xu, W.; Seung, S. J.; Mittmann, N.; Liu, G. EQ-5D Health Utility Scores: Data from a Comprehensive Canadian Cancer Centre. *Patient.* 2017;10(1):105-115.
17. Pickard ASJ, Ruixuan; Lin, Hsiang-Wen; Rosenbloom, Sarah; Cella, David. Using Patient-reported Outcomes to Compare Relative Burden of Cancer: EQ-5D and Functional Assessment of Cancer Therapy-General in Eleven Types of Cancer. *Clinical Therapeutics.* 2016;38(4):769-777.
18. Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2011;31(6):800-804.
19. Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. *Med Care.* 2005;43(7):736-749.

20. Cohen JW, Monheit AC, Beauregard KM, et al. The Medical Expenditure Panel Survey: a national health information resource. *Inquiry.* 1996;33(4):373-389.
21. Kulkarni GS, Alibhai SM, Finelli A, et al. Cost-effectiveness analysis of immediate radical cystectomy versus intravesical Bacillus Calmette-Guerin therapy for high-risk, high-grade (T1G3) bladder cancer. *Cancer.* 2009;115(23):5450-5459.
22. Xie F, Zoratti M, Chan K, et al. 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). *Medical decision making : an international journal of the Society for Medical Decision Making.* 2019;39(4):370-378.

Tables and Figures

Table 1: Study and respondent characteristics of health utility studies in bladder cancer

Table 2: Health states and utility values reported for bladder cancer

Figure 1: Summary of the study selection process

Figure 2: Health utility estimates for the bladder cancer health state

Table 1: Study and respondent characteristics of health utility studies in bladder cancer

Study	Scaling method(s)	Respondents	Number of respondents	Mean age (SD)*	Males, n (%)*	Country
Sullivan et al., 2011	EQ-5D-3L	General public, with some by proxy**	72 (CCC 032) 63 (ICD 188)	69.5 (CCC 032) 63.9 (ICD 188)	42,534 (48)	United States
Hays et al., 2014	SF-6D	Patients, with some by proxy	1299	75	12,052 (53)	United States
Kent et al., 2015	SF-6D	Patients	2035	70.1 (8.9)	2457 (76.9)	United States
Hever et al., 2015	EQ-5D-3L SF-6D	Patients	148**	66.3 (9.6)	98 (65)	Hungary
Pickard et al., 2016	EQ-5D-3L	Patients	31	62.5 (8.8)	276 (52)	United States
Naik et al., 2017	EQ-5D-3L	Patients	13	59*** (18, 100)	827 (47)	Canada

*With the exception of Hever et al., 2015 and Sullivan et al., 2011 (age only), summary measures are based on the whole study cohort and are not specific to patients with bladder cancer
**General public responding to the Medical Expenditure Panel Survey. Data is reported specifically for patients with bladder cancer per ICD and CCC codes.
***Number of respondents included in analysis varied by analytical approach
****Median (Minimum, Maximum)

Table 2: Health states and utility values reported for bladder cancer

Author	Health state	Scaling method	Analysis N	Mean (SD or 95% CI)	Country-specific weighting algorithm
Sullivan et al., 2011	Bladder cancer Per CCC 032: Cancer of bladder	EQ-5D-3L	72	0.71 (0.65, 0.77)	United Kingdom
	Bladder cancer Per ICD-9 188: Malignant neoplasm of bladder	EQ-5D-3L	63	0.71 (0.65, 0.78)	United Kingdom
Hays et al., 2014	Bladder cancer	SF-6D	1299	0.722 (0.003)	United Kingdom
				0.793 (0.003)	
Kent et al., 2015	Bladder cancer	SF-6D	2035	0.70 (0.69, 0.70)	United States [Assumed]
Hever et al., 2015	Bladder cancer	EQ-5D-3L	148	0.784 (0.242)	United Kingdom
		SF-6D (Bayesian v2, Posterior mean)	125	0.683 (0.136)	
		SF-6D (Bayesian, Parametric mean)	121	0.717 (0.141)	
		SF-6D (Ordinal v2, Standard gamble health state)	121	0.717 (0.141)	
		SF-6D (Ordinal v2, Ordinal health state)	125	0.738 (0.155)	
	Bladder cancer (45-54 years old)	EQ-5D-3L	--	0.751	
	Bladder cancer	EQ-5D-3L	--	0.794	

Author	Health state	Scaling method	Analysis N	Mean (SD or 95% CI)	Country-specific weighting algorithm
	(55-64 years old)				
	Bladder cancer (65-74 years old)	EQ-5D-3L	--	0.808	
	Bladder cancer (74-85 years)	EQ-5D-3L	--	0.728	
	Ileal conduit cystectomy	EQ-5D-3L	14	0.617 (0.354)	
	Ileal conduit cystectomy	SF-6D	14	0.623 (0.113)	
	Neobladder cystectomy	SF-6D	6	0.702 (0.193)	
	Neobladder cystectomy	EQ-5D-3L	6	0.81 (0.193)	
	Transurethral resection	SF-6D	63	0.739 (0.137)	
	Transurethral resection	EQ-5D-3L	63	0.788 (0.264)	
	Transurethral resection with intravesical therapy	SF-6D	68	0.72 (0.143)	
	Transurethral resection with intravesical therapy	EQ-5D-3L	68	0.815 (0.179)	
Pickard et al., 2016	Bladder cancer (Broadly)	EQ-5D-3L	31	0.81 (0.13)	United States
Naik et al., 2017	Bladder cancer (Broadly)	EQ-5D-3L	13	0.77 (0.14)	Canada
		EQ-5D-3L	13	0.74 (0.22)	United Kingdom
		EQ-5D-3L	13	0.80 (0.14)	United States

CCC: Clinical Classification Categories; ICD-9: International Classification of Diseases, Ninth Revision
Multiple entries from a single study reflect multiple analytical approaches. With the exception of Sullivan et al., 2011 (rounded to 2 decimals), values are reported as they appear in their original publications.

Figure 1: Summary of the study selection process

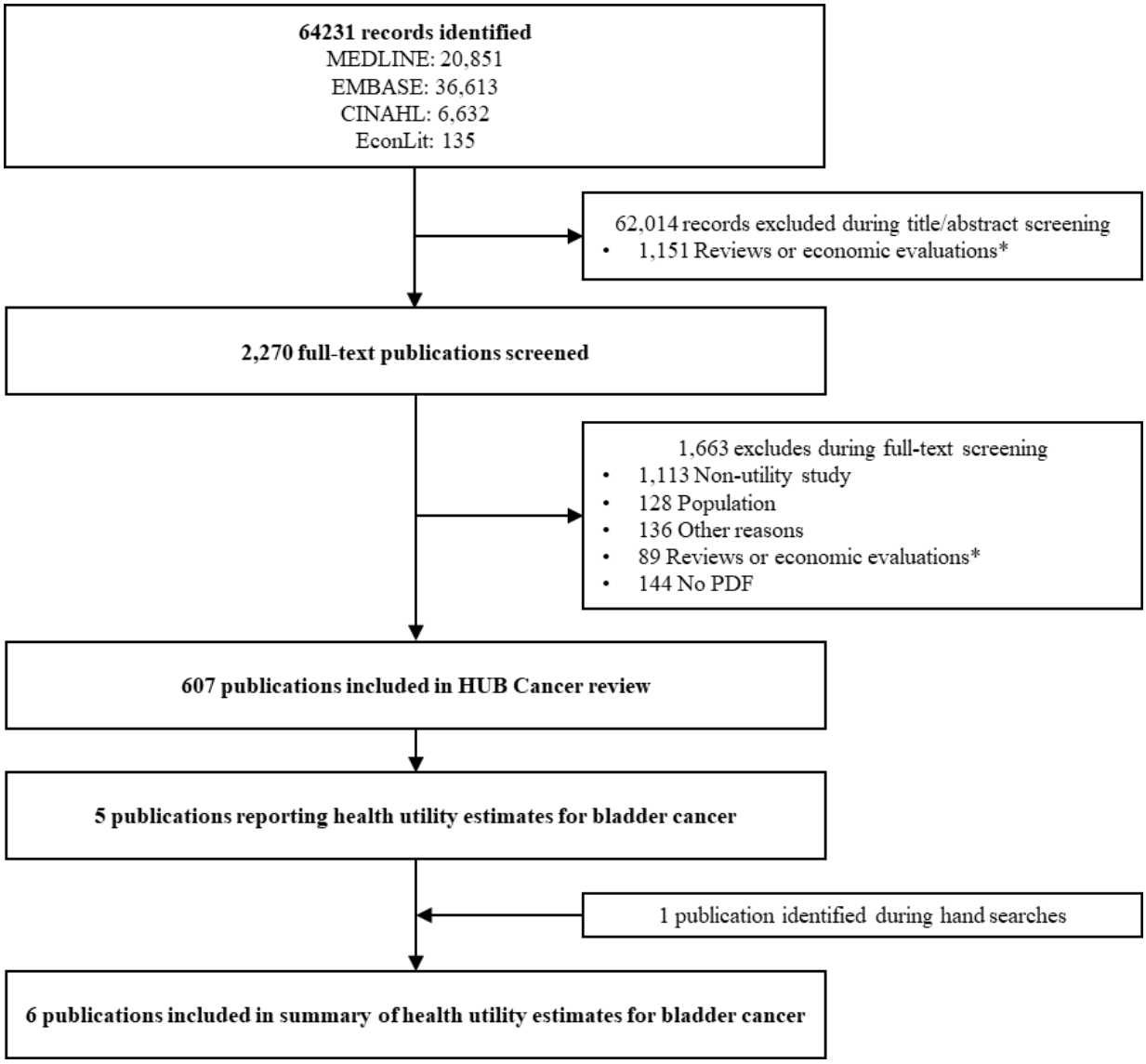
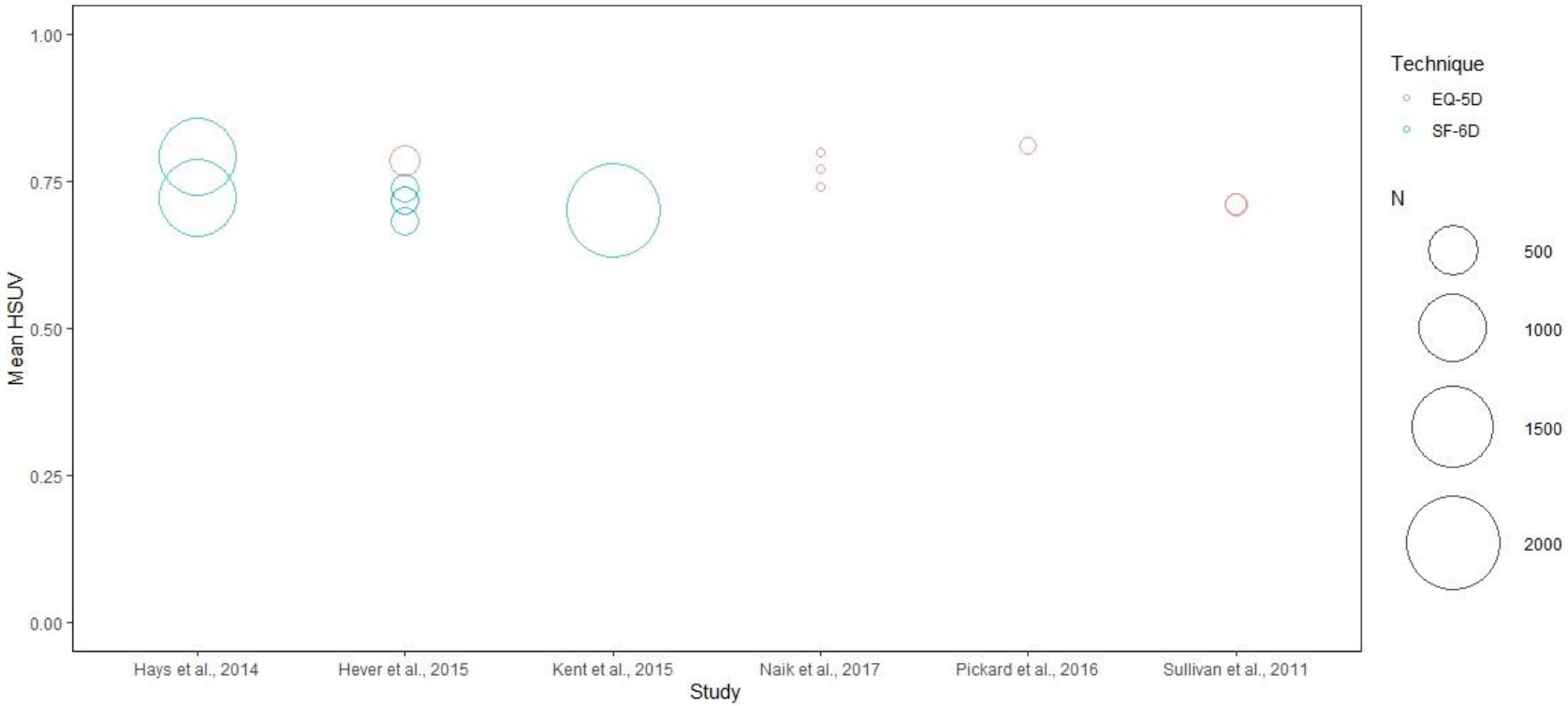


Figure 2: Health utility estimates for the bladder cancer health state



CHAPTER 3 – EVALUATING THE CONDUCT AND APPLICATION OF HEALTH UTILITY STUDIES: A REVIEW OF CRITICAL APPRAISAL TOOLS AND REPORTING CHECKLISTS

Status: This manuscript was published in the European Journal of Health Economics on April 11, 2021. It has been reproduced here with permission (License 5113680379037).

Original publication:

Zoratti MJ, Pickard AS, Stalmeier PFM, et al. Evaluating the conduct and application of health utility studies: a review of critical appraisal tools and reporting checklists. Eur J Health Econ. 2021;22(5):723-733.

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 Stalmeier,[3] Daniel Ollendorf,[4] Andrew Lloyd,[5] Kelvin Chan,[6,7] John E. Brazier,[8] Murray Krahn,[9] Mitchell Levine,[1,10] Lehana Thabane,[1,11] Feng Xie,[1] on behalf of the Health Utility Book (HUB) Working Group

Author affiliations:

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

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

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

[4]Value Measurement & Global Health Initiatives, Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center; Boston, Massachusetts, USA

[5]Acaster Lloyd Consulting Ltd; London, United Kingdom

[6]Sunnybrook Odette Cancer Centre; Toronto, Ontario, Canada

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

[8]School of Health and Related Research, University of Sheffield; Sheffield, United Kingdom

[9]Toronto General Hospital, Toronto, Ontario, Canada

[10]Department of Medicine, Division of Clinical Pharmacology and Toxicology, McMaster University; Hamilton, Ontario, Canada

[11]St Joseph's Healthcare - Hamilton; Hamilton, Ontario, Canada

Abstract

Background: Published health utility studies are increasingly cited in cost utility analysis 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.

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 -0.59 to +1.00 using the United Kingdom value set[12] while the SF-6Dv2 values range 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 judgement, 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, 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 is 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 modelled 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 10 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, 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 towards 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 (**Figure 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.

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 6 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 trade-off, 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 randomised 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 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 rigour in the application of the health utility literature in decision-making.

References

1. Klarman HE, Francis JOS, Rosenthal GD. *Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. [Article].* Medical Care January/February 1968;6(1):48-54.
2. Bremner KE, Chong CA, Tomlinson G, Alibhai SM, Krahn MD. A review and meta-analysis of prostate cancer utilities. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2007;27(3):288-298.
3. Sturza J. A review and meta-analysis of utility values for lung cancer. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2010;30(6):685-693.
4. Paracha N, Thuresson PO, Moreno SG, MacGilchrist KS. Health state utility values in locally advanced and metastatic breast cancer by treatment line: a systematic review. *Expert review of pharmacoeconomics & outcomes research.* 2016;16(5):549-559.
5. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. *Expert review of pharmacoeconomics & outcomes research.* 2010;10(5):553-566.
6. Hao Y, Wolfram V, Cook J. A structured review of health utility measures and elicitation in advanced/metastatic breast cancer. *ClinicoEconomics and outcomes research : CEOR.* 2016;8:293-303.
7. Schiller-Fruhwirth IC, Jahn B, Arvandi M, Siebert U. Cost-Effectiveness Models in Breast Cancer Screening in the General Population: A Systematic Review. *Applied health economics and health policy.* 2017;15(3):333-351.
8. Carter GC, King DT, Hess LM, et al. Health state utility values associated with advanced gastric, oesophageal, or gastro-oesophageal junction adenocarcinoma: a systematic review. *Journal of medical economics.* 2015;18(11):954-966.
9. Djalalov S, Rabeneck L, Tomlinson G, Bremner KE, Hilsden R, Hoch JS. A Review and Meta-analysis of Colorectal Cancer Utilities. *Medical decision*

making : an international journal of the Society for Medical Decision Making. 2014;34(6):809-818.

10. Jeong K, Cairns J. Systematic review of health state utility values for economic evaluation of colorectal cancer. *Health economics review.* 2016;6(1):36.
11. Richardson J, Khan MA, 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. *Medical decision making : an international journal of the Society for Medical Decision Making.* 2015;35(3):276-291.
12. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35(11):1095-1108.
13. Mulhern BJ, 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.* 2020;58(6):566-573.
14. Galante J, Augustovski F, Colantonio L, et al. Estimation and Comparison of EQ-5D Health States' Utility Weights for Pneumoccocal and Human Papillomavirus Diseases in Argentina, Chile, and the United Kingdom. *Value in Health.* 2011;14(5, Supplement):S60-S64.
15. Takemoto ML, Lopes da Silva N, Ribeiro-Pereira AC, Schilithz AO, Suzuki C. Differences in utility scores obtained through Brazilian and UK value sets: a cross-sectional study. *Health and quality of life outcomes.* 2015;13:119.
16. Pollard C, Hartz S, Leage SL, Paget MA, Cook J, Enstone A. Elicitation of health state utilities associated with varying severities of flares in Systemic Lupus Erythematosus. *Health and quality of life outcomes.* 2015;13:66.
17. Brazier J, Ara R, Azzabi I, et al. Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2019;22(3):267-275.

18. Xie F, Zoratti M, Chan K, et al. 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). *Medical decision making : an international journal of the Society for Medical Decision Making.* 2019;39(4):370-378.
19. Walker DG, Wilson RF, Sharma R, et al. *Best Practices for Conducting Economic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools.* Agency for Healthcare Research and Quality;2012.
20. Husereau D, Drummond M, Petrou S, et al. 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2013;16(2):231-250.
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. *International journal of technology assessment in health care.* 2005;21(2):240-245.
22. Nerich V, Saing S, Gamper EM, et al. Critical appraisal of health-state utility values used in breast cancer-related cost-utility analyses. *Breast cancer research and treatment.* 2017.
23. Chiou CF, Hay JW, Wallace JF, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care.* 2003;41(1):32-44.
24. Simoens S. Assessment of methodological quality of economic evaluations in belgian drug reimbursement applications. *PloS one.* 2013;8(12):e85411.
25. Stalmeier PF, Goldstein MK, Holmes AM, et al. What should be reported in a methods section on utility assessment? *Medical decision making : an international journal of the Society for Medical Decision Making.* 2001;21(3):200-207.

26. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. *PharmacoEconomics.* 2015;33(10):985-991.
27. Xie F, Pickard AS, Krabbe PF, et al. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). *PharmacoEconomics.* 2015;33(8):867-877.
28. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *Journal of health services research & policy.* 1999;4(3):174-184.
29. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ (Clinical research ed).* 1996;313(7052):275-283.
30. Ungar WJ, Santos MT. The Pediatric Quality Appraisal Questionnaire: an instrument for evaluation of the pediatric health economics literature. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2003;6(5):584-594.
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.* 1995;8(2):169-174.
32. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care.* 1992;30(3):231-243.
33. Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. *Health policy (Amsterdam, Netherlands).* 1992;21(3):249-279.
34. Sacristan JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. *The Annals of pharmacotherapy.* 1993;27(9):1126-1133.
35. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies.

International journal of technology assessment in health care. 2005;21(2):165-171.

36. Ramsey S, Willke R, Briggs A, et al. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA Task Force report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2005;8(5):521-533.
37. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Evidence and Value: Impact on DEcisionMaking--the EVIDEM framework and potential applications. *BMC Health Serv Res.* 2008;8:270.
38. Davis JC, Robertson MC, Comans T, Scuffham PA. Guidelines for conducting and reporting economic evaluation of fall prevention strategies. *Osteoporos Int.* 2011;22(9):2449-2459.
39. Vintzileos AM, Beazoglou T. Design, execution, interpretation, and reporting of economic evaluation studies in obstetrics. *Am J Obstet Gynecol.* 2004;191(4):1070-1076.
40. Grutters JP, Seferina SC, Tjan-Heijnen VC, van Kampen RJ, Goettsch WG, Joore MA. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14(5):777-784.
41. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *Jama.* 1996;276(14):1172-1177.
42. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *Jama.* 1996;276(16):1339-1341.
43. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *Jama.* 1996;276(15):1253-1258.

44. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *Jama.* 2016;316(10):1093-1103.

45. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical research ed).* 2011;343.

46. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539-549.

Tables and Figures

Table 1: Summary of 12 tools included in the review

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

Figure 1: Process of selecting tools and relevant items

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 1996	British Medical Journal (BMJ) Checklist	Reporting checklist	Yes/No	Expert consultation + Survey to achieve a broad consensus	35	2	2
Chiou 2003	Quality of Health Economic Studies (QHES) Instrument	Reporting checklist + Critical appraisal	Weighted scoring	Expert panel	16	1	1
Ungar 2003	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 1996	-	Reporting checklist	Yes/No	Expert panel	37	1	2
Husereau 2013	Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement	Reporting checklist	Indicate page numbers	Expert panel	24	1	1

Author/Year	Tool name	Primary purpose	Item response options	Development method	Number of items	Number of relevant items	
						Quality (n = 65)	Applicability (n = 44)
2nd US CEA Panel 2016	-	Reporting checklist	Yes/No	Expert panel	45	2	1
Tools specifically for health utility studies							
Brazier 1999	-	Critical appraisal	Open response	Author-proposed	24	3	5
Stalmeier 2001	-	Reporting checklist	Yes/No	Author-proposed + Expert panel	40	15	2
Petrou 2015	MApping onto Preference-based measures reporting Standards (MAPS) statement	Reporting checklist	Indicate page numbers	Expert panel	23	15	2
Xie 2015	Checklist for REporting VAluaTion StudiEs (CREATE)	Reporting checklist	Yes/No	Expert panel	21	13	3
Nerich 2017	-	Reporting checklist + Critical appraisal	Yes/No Open response	Author-proposed	17	4	12
Brazier 2019	ISPOR Minimum Reporting Standards of	Reporting checklist	Open response	Expert panel	15	5	10

Author/Year	Tool name	Primary purpose	Item response options	Development method	Number of items	Number of relevant items	
						Quality (n = 65)	Applicability (n = 44)
	Systematic Review of Utilities for Cost-Effectiveness models (ISPOR SpRUCE Checklist)						

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.

[1]Instruments where response options include a requirement to list the page number(s) corresponding to the item's criteria;
[2]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	✓	✓
Description of health states, if any	✓	✓
How is "perfect health" described?	✓	
How is "worst health" described?	✓	
If utility for "own health" was assessed, was own health specified further?	✓	
Description of treatments, if any (a treatment corresponds to a decision option in a decision tree)	✓	
Selection and description of respondents		
Whose quality of life is assessed?	✓	✓
Details of the subjects from whom evaluations were obtained are given	✓	✓
Is a comprehensive description provided for the population used to elicit HSUVs (i.e., characteristics, size, and nationality)?	✓	✓
Is an explanation provided for the choice of the population used to elicit HSUVs (i.e., patient, healthcare professional [and type], expert, general population)?	✓	✓
Response rate for the measure used	✓	
Extent of missing data or data lost to follow-up	✓	
Are the HSUVs appropriate with respect to comparability of populations (i.e., diagnosis and disease severity)?		✓
Are the HSUVs appropriate with respect to comparability of countries?		✓
Population or patient characteristics		✓
Are the items relevant and appropriate for the population?		✓
Whose values have been used?		✓
How well are the preferences of the patients/general population/decision-makers likely to conform to these assumptions?		✓
Are the background characteristics of the respondents to the valuation survey representative of the population?		✓

Item	Quality (n=65*)	Applicability (n=44*)
Measurement and valuation of preference based outcomes (If applicable, describe the population and methods used to elicit preferences for outcomes)	✓	✓
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	✓	✓
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)	✓	
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 summarise results of analyses used to assess overlap between the source and target measures	✓	
Sample size/power calculations are stated and rationalized	✓	
Target population is described	✓	✓
Sampling method is stated and rationalized	✓	
Recruitment strategies are described	✓	
Response rate is reported	✓	
Reasons for excluding any respondents or observations are provided	✓	
Characteristics of respondents included in the analysis are described	✓	✓
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?	✓	✓
Is an explanation provided for the choice of technique(s) used to elicit HSUVs?	✓	✓
If quality of life is measured, what type of instrument is used?	✓	✓
Methods to value health states and other benefits are stated	✓	✓

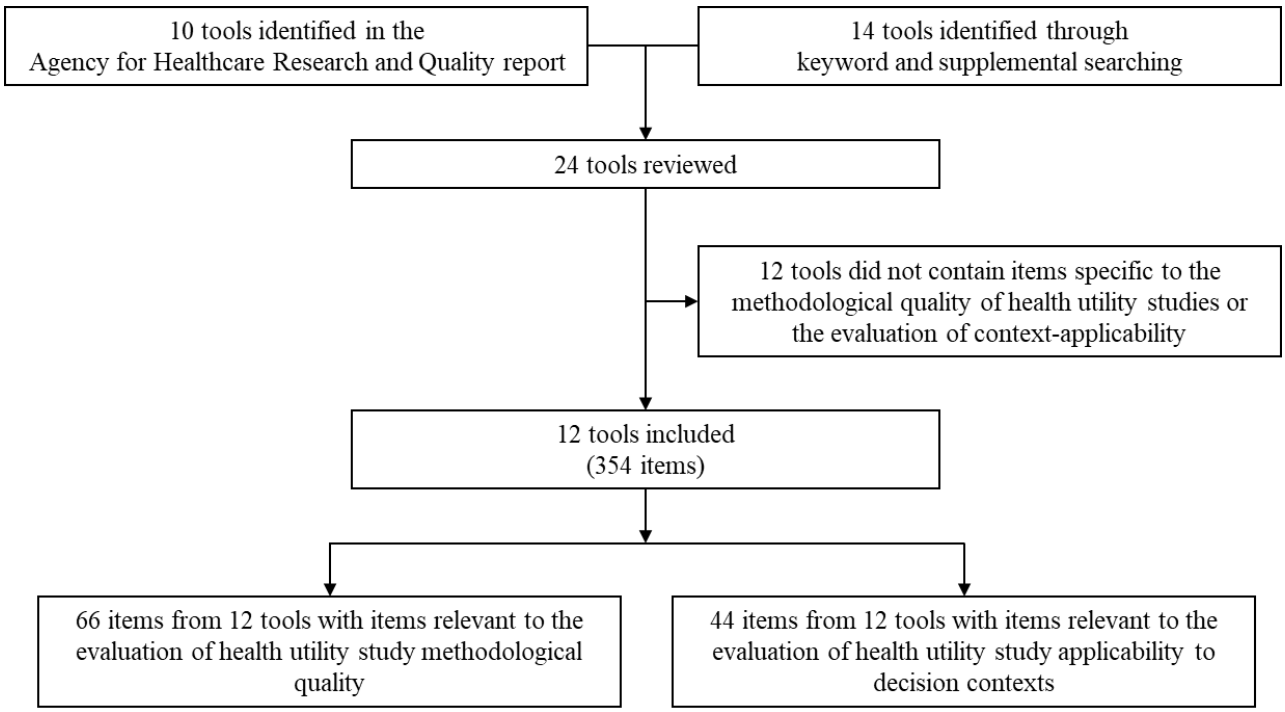
Item	Quality (n=65*)	Applicability (n=44*)
Is a comprehensive description provided of technique(s) used to elicit the obtained HSUVs?	✓	✓
If more than 1 utility measure was used, was the presentation order randomized? If not, what was the order?	✓	
Who performed the quality-of-life assessment?	✓	✓
How 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?	✓	
Was the presentation order of the health states randomized? If not, what was the order?	✓	
Was it made explicit that each duration was followed by death?	✓	
Was the subject instructed to assume that survival does not occur with knowledge of the date of death?	✓	
Was the health state labeled or unlabeled?	✓	
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?	✓	
Was a matching or choice indifference search procedure used? If yes, which particular indifference search procedure was used?	✓	
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?	✓	
What visual aids, if any, were used (e.g., rulers, pies, probability wheels, or other means of visualizing probabilities or trade-offs)?	✓	
What is the model of preferences being assumed?	✓	
What are the main assumptions of this model?	✓	
Which choice-based method has been used?	✓	
Measure used	✓	✓
Preference weights	✓	✓
Methods for obtaining estimates of effectiveness, costs and preferences		✓
Which method was chosen (e.g., visual analogue scale, time trade-off, standard gamble, willingness to pay)?		✓

Item	Quality (n=65*)	Applicability (n=44*)
Does the instrument cover all dimensions of health of interest?		✓
Descriptive statistics about HSUs		✓
Measurement and valuation of preference based outcomes (If applicable, describe the population and methods used to elicit preferences for outcomes)	✓	✓
Describe how predicted scores or utilities are estimated for each model specification	✓	
Methods for obtaining estimates of costs and preference weights	✓	✓
The attributes of the system are described	✓	
The number of levels in each attribute of the instrument is described	✓	
Method(s) of assigning the health states to respondents are stated	✓	
Mode of data collection is stated	✓	✓
Preference elicitation technique(s) are described	✓	
Other considerations		
Is the difference between when the CUA was performed and when the HSUVs were elicited less than 10 years?		✓
Do the authors use the same HSUVs in the CUA as presented in the original data source?		✓
Do the authors discuss the limitations of the data source selection, the elicitation, and the use of HSUVs in the CUA?		✓
Original reference		✓
Actual HSUs used		✓
Adjustments or assumptions		✓
Describe the source and target measures and the methods by which they were applied in the mapping study	✓	
Describe the methods used to assess the degree of conceptual overlap between the source and target measures	✓	
State how much data were missing and how missing data were handled in the sample(s) used for the analyses	✓	
Describe and justify the statistical model(s) used to develop the mapping algorithm	✓	
Describe and justify the methods used to validate the mapping algorithm	✓	

Item	Quality (n=65*)	Applicability (n=44*)
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	✓	
State which model(s) is(are) preferred and justify why this(these) model(s) was(were) chosen	✓	
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)	✓	
Report information that enables users to estimate standard errors around mean utility predictions and individual-level variability	✓	
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)	✓	
Critique of data quality	✓	
The approach to selecting health states to be valued directly is explained	✓	

Note: 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

Figure 1: Process of selecting tools and relevant items



Appendix A

This appendix presents each tool which included items that were considered to be of relevance to the concepts of critical appraisal and applicability. Relevant items are indicated for each tool.

Tools primarily intended for use with economic evaluations are presented in **Tables A1-A6**. Tools primarily intended for use with health utility studies are presented in **Tables A7-A12**.

Table A1: Items from the British Medical Journal Checklist by Drummond 1996

Item	Relevance	
	Quality	Applicability
Study design		
1. The research question is stated	×	×
2. The economic importance of the research question is stated	×	×
3. The viewpoints of the analysis are clearly stated and justified	×	×
4. The rationale for choosing the alternative programmes or interventions compared is stated	×	×
5. The alternatives being compared are clearly described	×	×
6. The form of economic evaluation used is stated	×	×
7. The choice of form of economic evaluation is justified in relation to the questions addressed	×	×
Data collection		
8. The sources of effectiveness estimates used are stated	×	×
9. Details of the design and results of effectiveness study are given (if based on a single study)	×	×
10. Details of the method of s changes to the study synthesis or meta-analysis of estimates are given (if base on an overview of a number of effectiveness studies)	×	×
11. The primary outcome measure(s) for the economic evaluation are clearly stated	×	×
12. Methods to value health states and other benefits are stated	✓	✓
13. Details of the subjects from whom evaluations were obtained are given	✓	✓
14. Productivity changes (if included) are reported separately	×	×
15. The relevance of productivity changes to the study question is discussed	×	×
16. Quantities of resources are reported separately from their unit costs	×	×
17. Methods for the estimation of quantities and unit costs are described	×	×
18. Currency and price data are recorded	×	×
19. Details of currency or price adjustments for inflation or currency conversion are given	×	×
20. Details of any model used are given	×	×
21. The choice of model used and the key parameters on which it is based are justified	×	×
Analysis and interpretation of results		
22. Time horizon of costs and benefits is stated	×	×
23. The discount rate(s) is stated	×	×
24. The choice of rate(s) is justified	×	×
25. An explanation is given if costs or benefits are not discounted	×	×
26. Details of statistical tests and confidence intervals are given for stochastic data	×	×
27. The approach to sensitivity analysis is given	×	×
28. The choice of variables for sensitivity analysis is justified	×	×
29. The ranges over which the variables are varied are stated	×	×
30. Relevant alternatives are compared	×	×
31. Incremental analysis is reported	×	×

Item	Relevance	
	Quality	Applicability
32. Major outcomes are presented in a disaggregated as well as aggregated form	×	×
33. The answer to the study question is given	×	×
34. Conclusion follow from the data reported	×	×
35. Conclusions are accompanied by the appropriate caveats	×	×

Table A2: Items from the Quality of Health Economic Studies (QHES) list by Chiou 2003

Item	Relevance	
	Quality	Applicability
1. Was the study objective presented in a clear, specific, and measurable manner?	×	×
2. Were the perspective of the analysis (societal, third-party payer, and so on) and reasons for its selection stated?	×	×
3. Were variable estimates used in the analysis from the best available source (i.e., Randomized Control Trial-Best, Expert Opinion-Worst)?	×	×
4. If estimates came from a subgroup analysis, were the groups pre-specified at the beginning of the study?	×	×
5. Was uncertainty handled by: (i) statistical analysis to address random events; (ii) sensitivity analysis to cover a range of assumptions?	×	×
6. Was incremental analysis performed between alternatives for resources and costs?	×	×
7. Was the methodology for data abstraction (including value health states and other benefits) stated?	×	×
8. Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3-5%) and justification given for the discount rate?	×	×
9. Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described?	×	×
10. Was the primary outcome measure(s) for the economic evaluation clearly stated and were the major short term, long term and negative outcomes included?	×	×
11. 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?	✓	✓
12. Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner?	×	×
13. Were the choices of economic model, main assumptions and limitations of the study stated and justified?	×	×
14. Did the author(s) explicitly discuss direction and magnitude of potential biases?	×	×
15. Were the conclusions/recommendations of the study justified and based on the study results?	×	×
16. Was there a statement disclosing the source of funding for the study?	×	×

Table A3: Items from the Pediatric Quality Appraisal Questionnaire (PQAQ) by Ungar 2003

Item	Relevance	
	Quality	Applicability
Economic Evaluation		
1. Is the research question posed in terms of costs and consequences?	×	×
2. Is a specific type of economic analysis technique performed?	×	×
3. What type of analytic technique is performed, according to the authors?	×	×
Comparators		
4. Is there a rationale for choosing the intervention(s) being investigated?	×	×
5. Is there a rationale for choosing the alternative program(s) or intervention(s) used for comparison?	×	×
6. Does the report describe the alternatives in adequate detail?	×	×
7. Is a description of the event pathway provided?	×	×
8. Is a formal decision analysis performed?	×	×
Target Population		
9. Is the target population for the intervention identified?	×	×
10. Are the subjects representative of the population to which the intervention is targeted?	×	×
Time Horizon		
11. Is there a time horizon for both costs and outcomes?	×	×
12. Do the authors justify the time horizon selected?	×	×
Perspective		
13. Is a perspective for the analysis given?	×	×
14. Is a societal perspective taken, either alone or in addition to other perspectives?	×	×
15. When there is more than one perspective, are the results of each perspective presented separately?	×	×
Costs and Resource Use		
16. Are all relevant costs for each alternative included?	×	×
17. Are opportunity costs of lost time (productivity costs) for parents and informal caregivers measured when required?	×	×
18. Do cost item identification and valuation extend beyond the health-care system to include school and community resources when necessary?	×	×
19. Are future salary and productivity changes of the child taken into consideration when appropriate?	×	×
20. Are all of the sources for estimating the volume of resource use described?	×	×
21. Are all the sources for estimating all of the unit costs described?	×	×
Outcomes		
22. Is a primary health outcome given?	×	×
23. Do the authors justify the health outcome(s) selected?	×	×
24. Is effectiveness, rather than efficacy, assessed?	×	×
25. What approach is used to assess the effectiveness/efficacy?	×	×

Item	Relevance	
	Quality	Applicability
26. Are the details of the design of the effectiveness/efficacy study(s) provided?	×	×
27. Are the results of the efficacy/effectiveness of alternatives reported?	×	×
28. Are school/day-care absences taken into consideration?	×	×
29. If intermediate outcome variables are used, are they linked by evidence or reference to the end benefit?	×	×
Quality of Life		
30. If quality of life is measured, what type of instrument is used?	✓	✓
31. Whose quality of life is assessed?	✓	✓
32. Who performed the quality-of-life assessment?	✓	✓
Analysis		
33. Are costs AND outcomes measured in units appropriate for the indicated analytic technique?	×	×
34. For prospective studies that use interviews, questionnaires, or surveys, how are data obtained in studies involving young children?	×	×
35. How are direct costs valued?	×	×
36. How are productivity costs valued?	×	×
37. Are costs valued appropriated?	×	×
38. Is the valuation of outcomes appropriate for the type of analysis?	×	×
39. What is the unit of analysis used for expressing the *final* results?	×	×
40. Are quantities of resources used reported separately from their unit costs?	×	×
41. Are the costs aggregated correctly?	×	×
42. Are details of statistical tests and confidence intervals given for stochastic data?	×	×
Discounting		
43. When required, are costs and consequences that occur over more than 1 year discounted to their present values?	×	×
44. If costs or benefits are not discounted when the time horizon exceeds 1 year, is an explanation provided?	×	×
Incremental Analysis		
45. Are incremental estimates of costs and outcomes presented?	×	×
46. Are the incremental estimated summarized as incremental ratios?	×	×
47. Are confidence intervals/limits calculated for incremental ratios or incremental estimates of costs and outcomes?	×	×
Sensitivity Analysis		
48. Are all important assumptions given?	×	×
49. Is a sensitivity analysis performed?	×	×
50. Do the authors justified the alternative values or ranges for sensitivity analysis?	×	×
51. What methods are used to assess uncertainty?	×	×
Conflict of Interest		
52. Does the article present the relationship with the sponsor of the study?	×	×

Item	Relevance	
	Quality	Applicability
53. Does the article indicate that the authors had independent control over the methods and right to publish?	×	×
Conclusions		
54. Is the answer to the study question provided?	×	×
55. Are the most important limitations of the study discussed?	×	×
56. Do the authors generalize the conclusions to other settings or patient/client groups?	×	×
57. Global impression of the quality of the article	×	×

Table A4: Items from the US Panel 1996

Item	Relevance	
	Quality	Applicability
Framework		
1. Background of the problem	×	×
2. General framing and design of the analysis	×	×
3. Target population for intervention	×	×
4. Other program descriptors (eg, care setting, model of delivery, timing of intervention)	×	×
5. Description of comparator programs	×	×
6. Boundaries of the analysis	×	×
7. Time horizon	×	×
8. Statement of the perspective of the analysis	×	×
Data and Methods		
9. Description of event pathway	×	×
10. Identification of outcomes of interest in analysis	×	×
11. Description of model used	×	×
12. Modeling assumptions	×	×
13. Diagram of event pathway	×	×
14. Software used	×	×
15. Complete description of estimates of effectiveness, resource use, unit costs, health states, and quality of life weights and their sources	✓	✓
16. Methods for obtaining estimates of effectiveness, costs and preferences	×	✓
17. Critique of data quality	×	×
18. Statement of year of costs	×	×
19. Statement of method used to adjust costs for inflation	×	×
20. Statement and type of currency	×	×
21. Source and methods for obtaining expert judgment	×	×
22. Statement of discount rates	×	×
Results		
23. Results of model validation	×	×
24. Reference case results (discounted at 3% and undiscounted): total costs and effectiveness, incremental costs and effectiveness, and incremental cost effectiveness ratios	×	×
25. Results of sensitivity analyses	×	×
26. Other estimates of uncertainty, if available	×	×
27. Aggregate cost and effectiveness information	×	×
28. Disaggregated results, as relevant	×	×
29. Secondary analyses using 5% discount rate	×	×
30. Other secondary analyses, as relevant	×	×
Discussion		
31. Summary of reference case results	×	×
32. Summary of sensitivity of results to assumptions and uncertainties in the analysis	×	×
33. Discussion of analysis assumptions having important ethical implications	×	×
34. Limitations of the study	×	×

Item	Relevance	
	Quality	Applicability
35. Relevance of study results for specific policy questions or decisions	×	×
36. Results of related cost-effectiveness analyses	×	×
37. Distributive implications of an intervention	×	×

Table A5: Items from the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement

Item	Relevance	
	Quality	Applicability
Title and abstract		
1. Title Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	×	×
2. Abstract Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	×	×
Introduction		
3. Background and objectives Provide an explicit statement of the broader context for the study Present the study question and its relevance for health policy or practice decisions.	×	×
Methods		
4. Target population and subgroups Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	×	×
5. Setting and location State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	×	×
6. Study perspective Describe the perspective of the study and relate this to the costs being evaluated.	×	×
7. Comparators Describe the interventions or strategies being compared and state why they were chosen.	×	×
8. Time horizon State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	×	×
9. Discount rate Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	×	×
10. Choice of health outcomes Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	×	×
11. Measurement of effectiveness 11a. *Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data. 11b. *Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	× ×	× ×
12. Measurement and valuation of preference based outcomes	✓	✓

If applicable, describe the population and methods used to elicit preferences for outcomes		
13. Estimating resources and costs 13a. *Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs. *13b. Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	× ×	× ×
14. Currency, price date, and conversion Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	×	×
15. Choice of model Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	×	×
16. Assumptions Describe all structural or other assumptions underpinning the decision-analytical model.	×	×
17. Analytical methods Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	×	×
Results		
18. Study parameters Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	×	×
19. Incremental costs and outcomes For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios	×	×
20. Characterising uncertainty *20a. Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	×	×

20b. Model-based economic evaluation: Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	×	×
21. Characterising heterogeneity If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	×	×
Discussion		
22. Study findings, limitations, generalisability, and current knowledge Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	×	×
Other		
23. Source of funding Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	×	×
24. Conflicts of interest Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	×	×

Table A6: Items from the Second Panel on Cost-Effectiveness in Health and Medicine

Item	Relevance	
	Quality	Applicability
Introduction		
Background of the problem	×	×
Study design and scope		
Objectives	×	×
Audience	×	×
Type of analysis	×	×
Target populations	×	×
Description of interventions and comparators (including no intervention, if applicable)	×	×
Other intervention descriptors (eg, care setting, model of delivery, intensity and timing of intervention)	×	×
Boundaries of the analysis; defining the scope or comprehensiveness of the study (eg, for a screening program, whether only a subset of many possible strategies are included; for a transmissible condition, the extent to which disease transmission is captured; for interventions with many possible delivery settings, whether only one or more settings are modeled)	×	×
Time horizon	×	×
Analytic perspectives (eg, reference case perspectives [health care sector, societal]; other perspectives such as employer or payer)	×	×
Whether this analysis meets the requirements of the reference case	×	×
Analysis plan	×	×
Methods and data		
Trial-based analysis or model-based analysis. If model based:	×	×
Description of	×	×
Description of event pathway or model (describe condition or disease and the health states included)	×	×
Diagram of event pathway or model (depicting the sequencing and possible transitions among the health states included)	×	×
Description of model used (eg, decision tree, state transition, microsimulation)	×	×
Modeling assumptions	×	×
Software used	×	×
Identification of key outcomes	×	×
Complete information on sources of effectiveness data, cost data, and preference weights	×	×
Methods for obtaining estimates of effectiveness (including approaches used for evidence synthesis)	×	×
Methods for obtaining estimates of costs and preference weights	✓	✓
Critique of data quality	✓	×
Statement of costing year (ie, the year to which all costs have been adjusted for the analysis; eg, 2016)	×	×
Statement of method used to adjust costs for inflation	×	×
Statement of type of currency	×	×
Source and methods for obtaining expert judgment if applicable	×	×
Statement of discount rates	×	×

Item	Relevance	
	Quality	Applicability
Impact inventory		
Full accounting of consequences within and outside the health care sector	×	×
Results		
Results of model validation	×	×
Reference case results (discounted and undiscounted): total costs and effectiveness, incremental costs and effectiveness, incremental cost-effectiveness ratios, measures of uncertainty	×	×
Disaggregated results for important categories of costs, outcomes, or both	×	×
Results of sensitivity analysis	×	×
Other estimates of uncertainty	×	×
Graphical representation of cost-effectiveness results	×	×
Graphical representation of uncertainty analyses	×	×
Aggregate cost and effectiveness information	×	×
Secondary analyses	×	×
Disclosures		
Statement of any potential conflicts of interest due to funding source, collaborations, or outside interests	×	×
Discussion		
Summary of reference case results	×	×
Summary of sensitivity of results to assumptions and uncertainties in the analysis	×	×
Discussion of the study results in the context of results of related cost-effective analyses	×	×
Discussion of ethical implications (eg, distributive implications relating to age, disability, or other characteristics of the population)	×	×
Limitations of the study	×	×
Relevance of study results to specific policy questions or decisions	×	×

Table A7: Items from Brazier 1999

Item	Relevance	
	Quality	Applicability
Practicality		
How long does the instrument take to complete?	×	×
What is the response rate to the instrument?	×	×
What is the rate of completion?	×	×
Reliability		
What is the test-re-test reliability?	×	×
What are the implications for sample size?	×	×
What is the inter-rater reliability?	×	×
What is the reliability between places of administration?	×	×
Validity – Description – Content validity		
Does the instrument cover all dimensions of health of interest?	×	✓
Do the items appear sensitive enough?	×	×
Validity – Description – Face validity		
Are the items relevant and appropriate for the population?	×	✓
Validity – Description – Construct validity		
Can the unscored classification of the instrument detect known or expected differences or changes in health?	×	×
Validity – Valuation		
Whose values have been used?	×	✓
Validity – Valuation – Assumptions about preferences		
What is the model of preferences being assumed?	✓	×
What are the main assumptions of this model?	✓	×
How well are the preferences of the patients/general population/decision-makers likely to conform to these assumptions?	×	✓
Validity – Valuation – Technique of valuation		
Is it choice-based?	×	×
Which choice-based method has been used?	✓	×
Validity – Valuation – Quality of data		
Are the background characteristics of the respondents to the valuation survey representative of the population?	×	✓
What was the degree of variation in the valuation survey?	×	×
Was there evidence of the respondents' understanding of the task?	×	×
What was the method of estimation (where relevant)?	×	×
Validity – Empirical		
Is there any evidence for the empirical validity of the instrument?	×	×
Revealed preferences?	×	×
Stated preferences?	×	×
Hypothesised preferences?	×	×

Table A8: Items from Stalmeier 2001

Item	Relevance	
	Quality	Applicability
Design		
A1. When exactly were utility measurements done, for example, before or after a medical treatment?	×	×
A2. The timing of utility assessment relative to other questionnaires	×	×
Administration		
B1. How were the utility questions administered (e.g., by interview, mailed questionnaires, computer, the Internet, or self-administered under general supervision)? If by interview:	–	–
a. What was the interview setting (e.g., face-to-face, by telephone, or in the hospital)?	×	×
b. Where were the utility measurements done (city)?	×	×
c. How were interviewers trained?	✓	×
d. How was between-interviewer reliability assessed?	×	×
e. What was the interview duration?	×	×
B2. Response and completion rates	×	×
B3. Efforts (if any) to increase response or completion rates	×	×
Health State Descriptions		
C1. Description of health states, if any	✓	✓
C2. How is "perfect health" described?	✓	×
C3. How is "worst health" described?	✓	×
C4. If utility for "own health" was assessed, was own health specified further?	✓	×
C5. Was the presentation order of the health states randomized? If not, what was the order?	✓	×
C6. Were fixed survival durations used for health states? a. If so, what duration(s) was used?	×	×
C7. How were the survival durations characterized (e.g., a survival outcome might be described as *x* years of survival followed by death or as a life expectancy of *x* years of survival)?	×	×
C8. How was the fixed survival duration hosen (e.g., from life tables of the general population or data from studies on a particular disease)?	×	×
C9. Was it made explicit that each duration was followed by death?	✓	×
C10. Was the subject instructed to assume that survival does not occur with knowledge of the date of death?	✓	×
C11. Description of treatments, if any (a treatment corresponds to a decision option in a decision tree).	✓	×
C12. What was the subject instructed to assume, if anything, regarding costs to him or her or family about the possible outcomes?	×	×
C13. Was the health state labeled or unlabeled?	✓	×
Description of the Utility Assessment Method		
D1. Which method was chosen (e.g., visual analogue scale, time trade-off, standard gamble, willingness to pay)?	×	✓
D2. If more than 1 utility measure was used, was the presentation order randomized? If not, what was the order?	✓	×

Item	Relevance	
	Quality	Applicability
D3. How was the choice introduced?	×	×
D4. If more than 1 health stated was assessed, were they rank ordered first?	×	×
a. If so, was death included in the ordering procedure	×	×
b. If there were states worse than dead, how were they handled?	×	×
D5. 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?	✓	×
Indifference Procedures		
E1. Was a matching or choice indifference search procedure used? In the case of choice,	✓	×
a. What were the first 2 choices?	×	×
b. Which particular indifference search procedure was used?	✓	×
c. What were the criteria for terminating the indifference search procedure (e.g., terminating when the range is narrowed to 10%)?	×	×
d. Did the subject give a final guess within the narrowed-down indifference range?	×	×
Visual Aids and Software Programs		
F1. Which software program (if any) was used? If a software program was used,	✓	×
a. Was it used by the subject alone, or was someone present in the start-up phase to answer questions or detect misconceptions?	✓	×
b. If someone was present, how was he or she trained?	✓	×
F2. What visual aids, if any, were used (e.g., rulers, pies, probability wheels, or other means of visualizing probabilities or trade-offs)?	✓	×
F3. Were any aspects of the interview controlled by computer?	×	×

Table A9: Items from the MApping onto Preference-based measures reporting Standards (MAPS)

Item	Relevance	
	Quality	Applicability
Title and abstract		
Identify the report as a study mapping between outcome measures. State the source measure(s) and generic, preference-based target measure(s) used in the study	×	×
Provide a structured abstract including, as applicable: objectives; methods, including data sources and their key characteristics, outcome measures used and estimation and validation strategies; results, including indicators of model performance; conclusions; and implications of key findings	×	×
Introduction		
Describe the rationale for the mapping study in the context of the broader evidence base	×	×
Specify the research question with reference to the source and target measures use and the disease or population context of the study	×	×
Methods		
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	✓	✓
Describe the source and target measures and the methods by which they were applied in the mapping study	✓	×
Describe the methods used to assess the degree of conceptual overlap between the source and target measures	✓	×
State how much data were missing and how missing data were handled in the sample(s) used for the analyses	✓	×
Describe and justify the statistical model(s) used to develop the mapping algorithm	✓	×
Describe how predicted scores or utilities are estimated for each model specification	✓	×
Describe and justify the methods used to validate the mapping algorithm	✓	×
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	✓	×
Results		
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)	✓	×
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 summarise results of analyses used to assess overlap between the source and target measures	✓	×

Item	Relevance	
	Quality	Applicability
State which model(s) is(are) preferred and justify why this(these) model(s) was(were) chosen	✓	×
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)	✓	×
Report information that enables users to estimate standard errors around mean utility predictions and individual-level variability	✓	×
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)	✓	×
Discussion		
Report details of previously published studies developing mapping algorithms between the same source and target measures and describe differences between the algorithms, in terms of model performance, predictions and coefficients, if applicable	×	×
Outline the potential limitations of the mapping algorithm	×	×
Outline the clinical and research settings in which the mapping algorithm could be used	×	×
Other		
Describe the source(s) of funding and non-monetary support for the study, and the role of the funder(s) in its design, conduct and report. Report any conflicts of interest surrounding the roles of authors and funders	×	×

Table A10: Items from the Checklist for REporting VAluaTion StudiEs (CREATE)

Item	Relevance	
	Quality	Applicability
Descriptive system		
The attributes of the instrument are described	✓	×
The number of levels in each attribute of the instrument is described	✓	×
Health states valued		
The approach to selecting health states to be valued directly is explained	✓	×
The number of health states valued per respondent is stated	×	×
Method(s) of assigning the health states to respondents are stated	✓	×
Sampling		
Sample size/power calculations are stated and rationalized	✓	×
Target population is described	✓	✓
Sampling method is stated and rationalized	✓	×
Recruitment strategies are described	✓	×
Response rate is reported	✓	×
Preference data collection		
Mode of data collection is stated	✓	✓
Preference elicitation technique(s) are described	✓	×
Study sample		
Reasons for excluding any respondents or observations are provided	✓	×
Characteristics of respondents included in the analysis are described	✓	✓
Modeling		
The dependent variable for each model is stated	×	×
Independent variables for each model are explained	×	×
Model specifications are provided	×	×
Model estimators are described	×	×
Goodness-of-fit statistics for each model are reported	×	×
Scoring algorithm		
Criteria for selecting the preferred model are stated	×	×
The scoring algorithm is presented	×	×

Table A11: Items from Nerich 2017

Item	Relevance	
	Quality	Applicability
Data source selection of HSUV used by authors of CUA		
S1. What is (are) the data source(s) of HSUVs?	×	✓
S2. If HSUVs are derived from the literature, how many references are given?	×	×
S3. If derived from the literature, what is (are) the data source(s) of HSUVs?	×	✓
Elicitation of HSUVs used by authors of CUA		
E1a. Is an explanation provided for the choice of technique(s) used to elicit HSUVs?	✓	✓
E1b. Is a comprehensive description provided of technique(s) used to elicit the obtained HSUVs?	✓	✓
E2a. Is an explanation provided for the choice of the population used to elicit HSUVs (i.e., patient, healthcare professional [and type], expert, general population)?	✓	✓
E2b. Is a comprehensive description provided for the population used to elicit HSUVs (i.e., characteristics, size, and nationality)?	✓	✓
Use of HSUVs by authors of CUAs		
U1. Are the HSUVs appropriate with respect to comparability of populations (i.e., diagnosis and disease severity)?	×	✓
U2. Are the HSUVs appropriate with respect to comparability of countries?	×	✓
U3. Is the difference between when the CUA was performed and when the HSUVs were elicited less than 10 years?	×	✓
U4. Do the authors use specific utility values for each health state of the model in the CUA?	×	×
U5. Do the authors use only a single source of utility values for each health state of the model in the CUA?	×	×
U6. Do the authors use specific utility values for each of the compared interventions in the CUA?	×	×
U7. Do the authors use the same HSUVs in the CUA as presented in the original data source?	×	✓
U8. Do the authors provide a comprehensive description and explanation for the explicit assumptions on the use of the HSUVs in the CUA?	×	✓
U9. Do the authors report results from a deterministic and probabilistic sensitivity analysis for the HSUVs?	×	×
U10. Do the authors discuss the limitations of the data source selection, the elicitation, and the use of HSUVs in the CUA?	×	✓

Table A12: Items from the ISPOR Minimum Reporting Standards of Systematic Review of Utilities for Cost-Effectiveness (SpRUCE) checklist

Item	Relevance	
	Quality	Applicability
Search strategy		
Search terms and scope	×	×
Study selection and scope	×	×
Review process		
Quality check	✓	✓
Assessment of HSU relevance	×	✓
Data extracted and reported		
Population or patient characteristics	×	✓
Measure used	✓	✓
Preference weights	✓	✓
Descriptive statistics about HSUs	×	✓
Response rate for the measure used	✓	×
Extent of missing data or data lost to follow-up	✓	×
Original reference	×	✓
Selection and estimation of final HSUs for the cost-effectiveness model		
Basis for selecting HSUs	×	✓
Method used to combine estimates	×	×
Methods used to apply the HSUs in the model		
Actual HSUs used	×	✓
Adjustments or assumptions	×	✓

CHAPTER 4 – TOWARDS TRANSPARENCY IN THE SELECTION OF PUBLISHED HEALTH UTILITY INPUTS IN COST-UTILITY ANALYSES: THE HEALTH UTILITY APPLICATION TOOL (HAT)

Status: This manuscript was published in PharmacoEconomics on June 21, 2021. It has been reproduced here with permission (License 5113680733458).

Original publication:

Zoratti MJ, Chan KKW, Husereau D, et al. Towards Transparency in the Selection of Published Health Utility Inputs in Cost-Utility Analyses: The Health Utility Application Tool (HAT). Pharmacoeconomics. 2021.

Towards transparency in the selection of published health utility inputs in cost-utility analyses: The Health utility Application Tool (HAT)

Michael J Zoratti,[1] Kelvin Chan,[2] Don Husereau,[3] Murray Krahn,[4] Mitchell Levine,[1,5] Lehana Thabane,[1,6] Feng Xie,[1,7,8] on behalf of the HUB Working Group

Author affiliations:

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

[2]Sunnybrook Odette Cancer Centre; Toronto, Ontario, Canada

[3]School of Epidemiology and Public health, University of Ottawa; Ottawa, Ontario, Canada

[4]Toronto General Hospital Research Institute, University Health Network; Toronto, Ontario, Canada

[5]Department of Medicine, Division of Clinical Pharmacology and Toxicology, McMaster University; Hamilton, Ontario, Canada

[6]St Joseph's Healthcare - Hamilton; Hamilton, Ontario, Canada

[7]Program for Health Economics and Outcome Measures (PHENOM); Hamilton, Ontario, Canada

[8]Centre for Health Economics and Policy Analysis (CHEPA), McMaster University; Hamilton, Ontario, Canada

Acknowledgements

We thank everyone who contributed to the development of this tool through either a telephone interview or response to our online survey. While all participants had the option of remaining anonymous, the following have agreed to be acknowledged:

Kabirraaj Toor, Whitney Longstaff, Colin Vicente, Lauren Cipriano, Donna Lawrence, Brittany Humphries, Wendy Ungar, A. Simon Pickard, Ahmed M.Y. Osman, Alex Li, Bernhard Michalowsky, Bin Wu, Boxiong Tang, Chang, Chao Wang, Charlene Niu, Chen, Chris Cameron, Ena Pery Niño de Guzman, Ernest Law, Gian Paolo Morgano, Gina, Graeme Ball, Haijing Guan, Hansen, Henry J. Conter, Hongchao Li, James W. Shaw, Jeffrey Johnson, Jie You, Jing Wu, Juan M Cabasés, Junwei Gao, Junwen Zhou, Lei Si, Lidia García, Marisa da Silva Santos, Matthijs Versteegh, Melina Leng, Nathan McClure, Nigar Sekercioglu, Olga Veljanoski, Paula Lorgelly, Peijing Yan, Rachel Harrington, Rosalie Viney, Shan Jiang, Sheng Wang, Sumeet Singh, Tania Conte, Tara Lavelle, Ting Zhou, Valentina Rupel, Weijin Fang, Wenying He, Xiangli Cui, Xiaoping, Xue Li, Yangwei, Yangyi, Zhiyuan Chen, and Huijun Zhou.

Abstract

Background: Evaluating the relevance of published health utilities to the context of a cost-utility analysis (CUA) remains an essential - yet often overlooked - task. The objective of this study was to provide guidance on this process through the development of the Health utility Application Tool (HAT).

Methods: We conducted semi-structured telephone interviews with Canadian stakeholders from reimbursement bodies, academia, and the pharmaceutical industry to identify current practices and perspectives of the application of the health utility literature to CUAs. An online survey with international members of the general health economics and outcomes research community was also conducted to gather opinions on key concepts.

Results: Based on the themes emerging from the interviews and online questionnaire, the HAT includes questions prompting investigators to consider the following constructs: Similarity of the clinical condition in the health utility study and the CUA; Similarity of health utility study participant demographics and the demographics of the CUA's target population; Similarity of the health state descriptions in the health utility study and the CUA; and the method of assigning utility weights. Considerations of transparency prompted additional items, including: Means by which the health utility study was identified; Type of respondents; Study design; and Measure used to collect health utility estimates.

Conclusion: The HAT is intended to guide the evaluation of the applicability of published health utilities for a CUA, thus promoting transparency and accountability in the selection of model inputs.

Introduction

Specifications for cost utility analyses (CUAs) are formally incorporated into the guidelines of several prominent health technology assessment (HTA) bodies, including those published by the National Institute for Health and Care Excellence (NICE) in the United Kingdom [1], the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia [2], and the Canadian Agency for Drugs and Technologies in Health (CADTH) [3]. However, while there is an intuitive understanding of the need to evaluate applicability and context relevance when sourcing input parameters from the literature, the guidance for the means by which health utilities are selected and incorporated into CUAs is varied and often limited to high-level recommendations. For instance, NICE guidance suggests that inputs should come from the "best available published sources" [1]. More specific guidance is provided by other HTA agencies such as the PBAC, where investigators are directed to describe "how representative the health state in each identified study is of the health state in the economic evaluation", with additional instructions to refer to the type and severity of symptoms and the duration of the health states [2]. Guidelines issued by CADTH stress the need for investigators to provide justification, broadly stating that the "best estimate, weighing trade-offs among the criteria of fitness for purpose, credibility, and consistency" should be selected for the reference case analysis [3]. When choosing "best" health utilities for a health economic evaluation, it is necessary to distinguish relevance from quality, although both are equally important to inform such a choice [4].

Evaluating the applicability or relevance of the available evidence is indeed a central consideration for investigators whenever the literature is to be applied to a

decision problem [5]. Evaluations of criteria such as population characteristics, choice of instruments, and means by which health states are valued have been proposed [6]. However, to our knowledge, no tool is available to guide investigators through these assessments. Here, we describe the development of the Health utility Application Tool (HAT), a tool designed to provide investigators with structured considerations and a means of making explicit their justifications for selecting among available health utility inputs.

Methods

The HAT was developed by building on existing guidance and engaging stakeholders and knowledge users. A literature review was conducted to identify how health utilities have been incorporated into published tools intended for use with the health economic literature (*publication pending*). The results of this review guided the development of our study materials.

We engaged stakeholders from academia, HTA bodies, pharmaceutical companies, and contract research organizations in Canada to participate in semi-structured telephone interviews to discuss methods and issues relating to the topic of applicability considerations when using published health utilities for CUAs. As part of our knowledge translation strategy, where we engage a breadth of stakeholders across the health economics and outcomes research spectrum, the sample size for this study was determined by data saturation [7,8]. A semi-structured interview guide was developed as a means of gathering focused, comparable qualitative data across multiple discussants (**Appendix A**) [9]. Participants were asked questions relating to their professional experience, their experiences using health utilities in their work, their opinions on the key

issues relating to applicability, and their opinions on the formatting and design considerations that make a tool user-friendly and helpful in practice. Following each interview, participants were invited to review a copy of the annotated interview transcript with instructions to make any addition, omission, or amendment they felt necessary. Anonymized interview transcripts were then reviewed and data were extracted through a thematic analysis to systematically identify common themes. Analyses were conducted independently and in duplicate by two investigators (MZ, FX). An inductive, rather than predetermined, coding approach was adopted to allow themes to emerge from the data in an exploratory, rather than explanatory, approach [10].

In parallel with the telephone interviews, we conducted an online survey to engage the broader international research community. This survey consisted of a series of closed and open response option questions asking respondents to indicate the importance of concepts to the transparent and methodologically rigorous selection of published health utilities (**Appendix B**). Survey questions were piloted with our research group to evaluate clarity and to define meaningful response options. An open invitation was made through social media channels and the investigative teams' affiliation mailing distribution lists to any researcher engaged in health economic research. Respondents had the option to skip any question and summary statistics were prepared based on the number of responses received for each item. Questions about the importance of issues relating to the use of published health utilities provided respondents with a 5-point scale, anchored at 1 for 'Not important at all' to 5 for 'Extremely important'. In our summary analyses, we collapsed scores of 4 and 5 to indicate agreement that an item is important, 1 and 2 to

indicate disagreement with an item's importance, and 3 to indicate indifference to an item's importance. Respondents were also provided with a free-text response box to indicate other issues that should be considered when selecting among published health utility inputs.

The findings from the interview and survey were compiled and evaluated irrespective of source or respondent to generate a list of main themes relevant to the assessment of health utility context relevance. Based on this list, an initial draft of the HAT was developed to reflect key themes in a simple and streamlined format. This draft was presented to subject matter experts with experience developing economic evaluations or appraisal frameworks who had not participated in a telephone interview or online survey. Telephone interview participants were also invited to review and provide feedback. Amendments were made and a final version of the HAT was prepared and approved by our research group.

Ethics approval for this study, including telephone interviews and the online survey, was granted by the Hamilton Integrated Research Ethics Board (HiREB ID: 2019-5722-GRA).

Results

Tool development

Telephone interviews

Interview invitations were distributed by email to 30 individuals and organizations. From June to October 2019, 13 participants from 11 organizations scheduled and completed an interview (**Table 1**). Participants were primarily employed

by a pharmaceutical company, a contract research organization, or held an academic position at a Canadian university. Nearly all participants (12/13, 92%) reported holding at least a master's degree.

All respondents agreed that the concept of applicability was an important and often overlooked consideration in the development of CUAs. Two major themes of relevance and transparency emerged through the thematic analysis (**Table 2**). Each theme was conceptualized through various sub-themes which referred to specific elements of transparency and relevance that respondents considered particularly important in the context of developing economic evaluations. For instance, the theme of relevance was further conceptualized through several sub-themes referring to aspects such as disease characteristics and the representativeness of the samples from which health utility values were estimated. Specifically, it was recommended that investigators should consider whether the characteristics of those from whom health utilities are originally measured are representative of the target population and decision problem. State-specific utility values were largely preferred, with an emphasis that health state descriptions should reflect the disease characteristics described in the CUA. With respect to transparency, respondents were primarily concerned with the traceability of estimates to facilitate a complete critical evaluation of the original sources of health utility evidence. Approximately half of the respondents identified clinical trials as an ideal source of health utility evidence (6/13, 46%), with registries or catalogues, systematic reviews, and standalone health utility studies also being recommended.

Online survey

Conducted in parallel with the telephone interviews, the online survey received an initial 152 responses from June to October 2019. However, we excluded the responses from 18 respondents who did not complete the survey beyond the demographic questions. Thus, the summary of findings is based on 134 respondents (**Table 1**). Approximately half of the respondents had earned a doctorate (n = 74/134, 55.2%), worked in academia (n = 72, 53.7%), and self-identified as a health economist (n = 78, 58.2%). Most participants reported having ten or fewer years of experience in the field of health economics and outcomes research (n = 107, 79.9%).

The vast majority of respondents indicated that a rationale should be provided when selecting from multiple published health utility studies, based on a response of 4 or 5 on the 5-point scale (n = 119/132, 90.1%) and that a justification should be made when combining multiple health utilities to obtain a pooled estimate (n = 114/132, 86.3%). Similarly, respondents agreed that a justification should be provided when using health utility estimates from clinical areas other than the one evaluated in the economic model, such as in rare diseases (n = 111/132, 84.1%). Participants largely agreed that the health state descriptions used in collecting health utility estimates should reflect important characteristics of the model's health states (n = 117/134, 87.3%), that published health utilities should be identified through systematic reviews (n = 106/133, 79.7%), and that the economic evaluation's target patient population should be the same as that from which health utilities were obtained (n = 94/132, 71.2%). In the context of a reimbursement submission, 73.7% of respondents (98/133) indicated that it is important that the methods used to measure health utilities are aligned with decision-maker

recommendations. Similarly, respondents agreed that utility-based instruments should utilize preference-based scoring algorithms specific to the country in which the economic evaluation is based (104/132, 78.8%). Respondents were indifferent to whether time since publication is an important consideration in the selection of health utilities.

Few differences in responses were observed when responses were stratified by participant characteristics, such as education and years of experience. However, compared to respondents with a master's degree, a higher proportion of individuals with a doctorate indicated that it is important for health state descriptions to match the health states in the economic model being developed and that investigators should measure health utility values according to the requirements and guidelines published by jurisdictional decision makers. Similarly, respondents with less than five years of experience were more likely than those with more than five years of experience to indicate that a systematic literature review is important to guide the identification of health utility inputs and that more recent health utility data is preferred.

The Health utility Application Tool

The HAT (**Box 1**) is defined by two main sections, one comprising a list of reporting criteria (Study Vitals) and the other consisting of contextual review questions (Context Assessment). Study Vitals refers to the characteristics of the source study where health utilities were originally estimated, including citation details, the original health utility estimates as reported in the source study, and the means by which the evidence was identified. Investigators should also identify the study design, type of respondents sampled, and the measure used to collect health utility estimates. Identifying these

elements allows for comparisons across the health utility inputs in an economic evaluation with respect to where, how, and from whom health utility inputs are being gathered.

Beyond the basic characteristics of the source of health utilities, we have proposed four additional questions to evaluate applicability (Context Assessment). Three of these questions direct investigators to indicate the degree of similarity between the health utility source and the model context based on three response options: very similar, somewhat similar, and different. An explanation should be provided to justify response selections and the potential implications of using this evidence, particularly when selecting either of the latter two options.

1) Is the clinical condition of the patients described in the health states the same as the clinical condition in the economic evaluation?

It is generally recommended to select health utility literature based on the clinical condition evaluated in the CUA. Doing so lends some degree of applicability or relevance simply from a common diagnosis and anticipated clinical experience. However, it is not uncommon to adopt health utility estimates from clinical areas that share similar characteristics and levels of disability to the condition evaluated in the CUA, particularly in the context of rare diseases where the literature simply does not exist. In such cases, judgments on clinical ambivalence should be made explicit and defendable.

2) Are the socio-demographics of participants in the health utility study the same as the socio-demographics of the target population for the economic evaluation?

Economic evaluations are based on specified perspectives and settings which drive the selection and estimation of costing data. This may also, to some extent, drive the selection of clinical model inputs to reflect patient outcomes in a select population. In this sense, the selection of health utility inputs should take into account the socio-demographics of the defined context.

3) Is the health state description in the health utility study the same as the health state description in the economic evaluation?

Consideration should be given as to whether the characteristics that define the health state as described in the health utility study reflect the characteristics modelled in the CUA. Of particular importance are the characteristics that drive health status, such as disease severity or functional ability. This applies to both vignette-based health utility exercises, where investigators present respondents with detailed descriptions of health status, as well as studies in which respondents complete assessments based on their own health using either direct or indirect measures.

4) For health utilities derived from utility-based instruments, which value set was used?

Value sets refer to collections of utility values for all possible health states that a utility-based instrument may describe. These preference-based weights are intended to reflect society's valuation of health states. However, the scoring functions used to generate value sets result in variations in health utility estimates for the same health state across different countries, though the reasons for these variations are not well understood.[11] Therefore, where available, a value set specific to the perspective country is typically recommended. In cases where a country-specific value set is not available, a

judgment is necessary to select estimates derived using a value set from a country or region that approximates the target setting.

As the characteristics of health utilities are unique to each health state, investigators should similarly complete the HAT for each health state utility input in the economic evaluation.

Discussion

Drawing from a review of the literature, engagement with stakeholders and anticipated end-users through telephone interviews and an online survey, and the collective experience of our research group, we have proposed the HAT as a tool that promotes transparency in the selection of published health utility values for CUAs by encouraging investigators to document justifications and reasonings in a transparent and structured manner. This is of particular importance for investigators working in the context of coverage or reimbursement decision making where funding allocations may be decided, in part, by the results of these evaluations. The tool is grounded in good practices, drawing upon constructs identified both in the literature and through engagement with knowledge users. Respondents to the online survey overwhelmingly indicated agreement with a statement concerning the need to provide justifications when selecting a health utility among a range of published values. In the absence of viable alternative tools, respondents' agreement with this statement establishes the value of the proposed HAT as a worthwhile and important development in economic evaluation methodology.

The HAT is intended to be used to evaluate each health utility parameter when developing an economic model. Its content is purposefully broad such that it may be used to evaluate the applicability of any health utility input, including estimates for health states, disutilities, or spillover effects. In cases where multiple health utilities are synthesized, for instance through a meta-analysis to generate a single estimate, we recommend investigators complete one HAT per evidence source. This applies also to pooled estimates that have been identified in the literature. Given the inherent subjectivity of the concepts which the HAT addresses, we propose this tool as a guide, rather than as a prescriptive set of rules by which investigators should abide. Judgements made with the HAT should be context-specific as, while citing previous considerations may support assessments, it is important to avoid simply re-applying evidence because it was considered acceptable in another application. Similarly, as the true value of this tool is found in the explanations provided by investigators for the decisions they make, the HAT does not include a summary score or attempt to visualize findings. By completing the HAT, we anticipate investigators will be able to better evaluate the appropriateness of the health utility literature they are sourcing for their models by making their implicit judgements explicit. To this end, the HAT serves as a standardized template which communicates these decisions and justifications to decision makers or other investigators who may adopt or otherwise use the CUA. The responsibility rests with the original investigator to provide appropriate rationales for their model inputs. However, while it may guide a reviewer to identify limitations or areas of uncertainty in a CUA, the HAT is not intended to be used for critical appraisal or evaluation. Instead, the HAT should be

used alongside a validated quality assessment tool. Reporting checklists and critical appraisal tools in the field of health economics vary in purpose yet share a common objective of improving transparency and promoting methods excellence. The HAT is not unique in this regard. In essence, it represents another attempt to promote good practices related to heath utilities. However, the focus on evaluating context applicability has not been, to our knowledge, the primary focus of any published health economic tool to date.

The strength of the HAT lies in its value to provide investigators with a clear and explicit outline of concepts to consider and address when defending their selection of health utility inputs in the development of a CUA. We proposed this tool through an iterative process with several data collection exercises, including an initial literature review and engagement of subject matter experts and knowledge users in multiple formats. Beyond the reporting items, we purposefully designed a tool that was lean and focused on the key constructs which our respondents felt would drive a health utility study's context applicability. Although efforts were made to recruit a variety of study participants to represent the collective users and doers of health utility research, we did have an overrepresentation of participants from the pharmaceutical industry and contract research organizations. However, as our research was based on methodologies, we are confident that the responses we received support our project objectives. Moreover, we did not observe differences in the types of responses and issues raised by either group of participants and many highlighted problems with the status quo approach to selecting health utilities. As a major source of the "doers" of health economic evaluations, engagement from the pharmaceutical industry is a strength of this work. Additionally, we

acknowledge that this tool was largely developed in a Canadian context with Canadian respondents. Further considerations may be necessary to adapt the HAT to other contexts, such as in cases where a specific approach to measuring health utilities is preferred by decision makers. However, the concepts represented are readily applicable to any decision-making jurisdiction as they were selected to represent a robust and transparent approach to selecting model inputs. Given the evolving nature of methods research, including methodological advancements in the fields of health economic evaluations, we do not anticipate the present version of the HAT to persist unchanged. Further research and application of the HAT may reveal additional constructs that should be included, and the tool should be reviewed and updated over time to reflect new methods standards. For example, while not highlighted as a key area of consideration by most of our study participants, time since publication may be proven an important factor when selecting among the most relevant evidence. The study group maintains that a useful tool should be flexible to emerging methods and the evolving needs of investigators.

Health technology assessment bodies provide generally high-level guidance for investigators on the identification and selection of health utility inputs for economic evaluations. This is likely a reflection of the lack of specific data available to investigators who must instead rely on evidence that they consider to be sufficiently similar to inform their model parameters. In light of imperfect evidence, we maintain that it is even more important for reasonings and justifications to be documented and made explicit to readers and reviewers. The use of the HAT promotes this position and encourages the appropriate and transparent application of health utilities in CUAs. We

anticipate the HAT being applied in all health economic research settings, including academia, the pharmaceutical industry, and HTA bodies and that its application will improve the credibility and relevance of CUAs to their specific decision contexts.

Conclusion

The Health Utility Book (HUB) project was initiated to promote selections of health utilities from the literature that are both credible, referring to a degree of methodological rigour, and relevant, referring to whether the estimates are fit for the context [4]. This multi-institutional collaboration is motivated by a need to provide investigators and decision makers with the guidance and tools to develop CUAs that are robust, transparent, and accountable. As a central objective of the HUB project, the HAT was developed to provide investigators with a means of documenting the processes through which published health utility inputs are selected. We achieved this through a review of existing tools and frameworks applied to the health economics literature and engagement with stakeholders and anticipated end-users. The final tool represents the collective body of good practices identified through these exercises based on our current understanding of methods and practices. Future iterations of the HAT may evolve from this groundwork. We hope that the HAT encourages transparent reporting of the selection of health utility inputs and that these selections are made through grounded principles.

References

1. National Institute for Health and Care Excellence. *Developing NICE guidelines: the manual*. 2014.
2. Australian Government. *Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0).* Sydney: Department of Health;2016.
3. CADTH. *Guidelines for the economic evaluation of health technologies: Canada. 4th ed.* Ottawa: CADTH; March 2017 2017.
4. Xie F, Zoratti M, Chan K, et al. 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). *Medical decision making : an international journal of the Society for Medical Decision Making.* 2019;39(4):370-378.
5. Ara R, Peasgood T, Mukuria C, et al. Sourcing and Using Appropriate Health State Utility Values in Economic Models in Health Care. *PharmacoEconomics.* 2017;35(1):7-9.
6. Papaioannou D, Brazier J, Paisley S. Systematic searching and selection of health state utility values from the literature. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2013;16(4):686-695.
7. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant.* 2018;52(4):1893-1907.
8. Burmeister E, Aitken LM. Sample size: how many is enough? *Aust Crit Care.* 2012;25(4):271-274.
9. Cohen D, Crabtree B. Qualitative Research Guidelines Project. http://www.qualres.org/HomeSemi-3629.html. Published 2006. Accessed July 10, 2017.
10. Guest G MK, Namey EE. *Applied Thematic Analysis.* Thousand Oaks: Sage; 2011.

11. Roudijk B, Donders ART, Stalmeier PFM. Cultural Values: Can They Explain Differences in Health Utilities between Countries? *Medical Decision Making.* 2019;39(5):605-616.

Tables and Boxes

Table 1: Interview and online survey participants

Table 2: Results of thematic analysis

Box 1: The Health utility Application Tool (HAT)

Table 1: Interview and online survey participants

Characteristic		Survey participants n = 134 n (%)	Interview participants n = 13 n (%)
Gender	Male	67 (50.8)	4 (30.8)
	Female	64 (48.5)	9 (69.2)
	Prefer not to say	1 (0.8)	–
Age (years)	18 – 30	33 (25)	2 (15.4)
	31 – 40	54 (40.9)	6 (46.2)
	41 – 50	28 (21.2)	4 (30.8)
	51 – 64	15 (11.4)	1 (7.7)
	65+	2 (1.5)	–
Highest level of education	Bachelor's degree	3 (2.1)	2 (15.4)
	Master's degree	51 (35.4)	7 (53.8)
	PhD	74 (51.4)	4 (30.8)
	Doctor of Medicine[1]	16 (11.1)	–
Affiliation[2]	Academia	72 (50)	2 (15.4)
	Government or non-profit organization	25 (17.4)	1 (7.7)
	Pharmaceutical company	17 (11.8)	2 (15.4)
	Consultancy	17 (11.8)	8 (61.5)
	Other	13 (9)	–
Years of experience in health economics and outcomes research	Less than 5 years	46 (34.3)	1 (7.7)
	5 – 10 years	11 (8.2)	5 (38.5)
	11 – 15 years	53 (39.6)	4 (30.8)
	More than 15 years	24 (17.9)	3 (23.1)
Approximate number of drug reimbursement submissions encountered in a typical year	None	49 (36.6)	1 (7.7)
	1 – 5	49 (36.6)	6 (46.2)
	6 – 19	20 (14.9)	5 (38.5)
	At least 20	16 (11.9)	1 (7.7)

Denominators for percentages reflect the number of respondents who provided a response.
[1]Number of respondents with a medical degree counted separately from other degrees;
[2]Respondents could indicate more than one affiliation.

Table 2: Results of the thematic analysis

Major theme	Sub-theme	Supporting quotes
Relevance of the health utility study to the decision problem	Similarity of disease characteristics between respondents and the decision problem	"Are diseases considered similar according to expert opinion?" A001 "Important to match the disease severities targeted by the economic evaluation with the persons from whom evidence is being collected from" C001 "When borrowing from other disease areas, consider the impact due to adverse events that may be similar across the diseases, populations" P001
	Similarity of respondent demographics to the decision problem	"Representativeness of the population versus the participants recruited in a waiting room" A002 "Appropriate to the country…there may be differences across different jurisdictions" C005 "Population from which utility estimates were collected reflects the target population of the cost utility analysis" C006
	Similarity of health state descriptions in the study where health utilities are measured and the decision problem	"Disease states…don't want to be using a utility score that doesn't match what you're trying to model" P002
Transparency	Original source of health utility estimates	"Need to cite the original source – stop citing other economic evaluations! We have some pretty persistent utility values in the literature." A002
	Means by which evidence was identified	"Oftentimes, when you start looking through the literature you identify a single source where all the health utility estimates come from. And this may not be the most well-done study." C004
	How health utility estimates were measured	"Difficult to determine how utilities were measured. For example, what tool was used?" C006

A00X refers to a participant from academia, P00X refers to a participant from a pharmaceutical company, and C00X refers to a participant from a contract research organization.

Box 1: The Health utility Application Tool

Complete one HAT per health utility input.

The HAT is intended to provide investigators with a summary of the key concepts to consider when selecting health utility inputs for cost-utility analyses. Investigators are encouraged to document their reasoning. It is not intended as a means of appraising or scoring selections.

Health state: Name of health state
Mean (SD) utility or disutility as reported: Mean (SD) HSUV
n: Number of respondents

Study Vitals
Describe the original health utility study.

Author	Publication author(s)
Year	Year of publication
Title	Publication title

Was this study identified through a systematic literature review or catalogue/registry?
☐ No ☐ Catalogue or registry ☐ Systematic review

If a systematic review, has it been published? ☐ Yes ☐ No
If yes, when was the review conducted? Year

Study design
☐ Randomized clinical trial
☐ Non-randomized clinical trial
☐ Observational study
☐ Meta-analysis of multiple studies
☐ Not clear

Type of respondents
☐ Patients ☐ Clinicians
☐ General public ☐ Caregiver or family members
☐ Other ☐ Not clear

What type of measure was used to collect health utility estimates?

☐ Direct measure (e.g., Standard gamble; Time trade-off)
☐ Utility-based instrument (e.g., EQ-5D, SF-6D)
☐ Mapping exercise from a disease-specific measure

Context Assessment

Assess the relevance of the health utility to the context of the economic evaluation. For each item, indicate the most correct response option and explain the selection and its implications.

1) Is the clinical condition of the patients described in the health states the same as the clinical condition in the economic evaluation?

☐ Very similar ☐ Somewhat similar ☐ Different

Explain your selection.

2) Are the socio-demographics of participants in the health utility study the same as the socio-demographics of the target population for the economic evaluation?

☐ Very similar ☐ Somewhat similar ☐ Different

Explain your selection.

3) Is the health state description in the health utility study the same as the health state description in the economic evaluation?

☐ Very similar ☐ Somewhat similar ☐ Different

Explain your selection.

4) For health utilities derived from utility-based instruments, which value set was used?

Value set: Specify the value set here.

If different from the decision context, explain how this applies.

Appendix A: Telephone interview script

Health utility Application Tool (HAT) – Interiew script

Section 1: Demographics and professional experience

Rationale: The items in this section will be used primarily to describe the demographic and professional characteristics of our respondent pool.

Demographics

1.1 Gender ☐ Male ☐ Female ☐ Other ☐ Prefer not to say

1.2 Age range ☐ 18-30 ☐ 31-40 ☐ 41-50 ☐ 51-64 ☐ 65+
☐ Prefer not to say

1.3 Education ☐ Bachelor's degree ☐ Master's degree
☐ Doctorate (Ph.D.) ☐ Medical degree (M.D.)
☐ Other degree: _____

1.4 What year did you complete your most recent degree? _____

Professional experience

1.5 Employer: _____

1.6 Current job title: _____

1.7 Years in current role
☐ <5 years ☐ 5-10 years ☐ 11-15 years ☐ More than 15 years

1.8 Years working in HEOR
☐ <5 years ☐ 5-10 years ☐ 11-15 years ☐ More than 15 years

Section 2: Experiences with economic evaluations

Rationale: This section describes the respondent pool with respect to content-area experience and will help contextualize participant responses.

2.1.a *For government and/or reviewing agencies (i.e. User):*

How many drug reimbursement submissions do you encounter in a given year? What is your role in that encounter?

Approximate Number:

Role:

2.1.b *For pharmaceutical industry (i.e. Doer):*

How many drug reimbursement submissions do you prepare or contribute to the preparation of in a given year?

Approximate Number:

2.2 *Use the following as follow-up probing questions:*

Have you ever been involved in the conduct of economic evaluations? This may include those encountered through your primary role or through other academic, consulting, or other ventures.

Yes No

2.3 *If yes, proceed with:*

In your work, do you contribute to the development of economic models, such as by providing guidance or recommendations on the selection model inputs? Are you primarily involved in model development or economic evaluations conducted alongside trials?

Yes No

Model-based Trial-based

[Comments]

Section 3: Use of health utilities

Rationale: Understand the processes used by the respondents and their organizations to evaluate health utilities, particularly with respect to their application to decision-making problems.
Understand the respondents' perception of the importance of evaluating applicability and their perception of adopting a formal process to address this need.

3.1 Are you familiar with health utilities?
Yes No

3.2 If you develop economic models, do you use health utility inputs?
Yes No Not applicable

3.3 Are you involved in studies or exercises where health utilities are measured?
Yes No

3.4 When reviewing (for government) or preparing (for industry) economic evaluations, do you find issues with the health utility inputs?
[Comments]

3.5 *Use the following as follow-up probing questions:*

3.5.a Do you often find that there is no evidence or only low-quality evidence available for disease areas or specific health states?
[Comments]

3.5.b Is there a lack of Canadian health utilities?
[Comments]

3.5.c Do you find that multiple eligible health utilities have been identified which makes it hard to choose one?
[Comments]

3.5.d Do you pool health utilities from multiple sources to get an average estimate?
Yes No
[Comments]

3.6.e What are some of the common justifications you encounter and/or provide for choosing among existing health utilities?
[Comments]

3.5.f In your opinion and according to your experience, what health utilities are ideal to be used in Canadian reimbursement submissions? This can include estimates identified through a systematic review, collected alongside trials, or elicited in stand-alone studies.
[Comments]

Section 4: Sample questions

Rationale: Collect feedback on sample question content, formatting, and response options.

Presented here is a series of questions. Each set presents the same content (i.e. construct), though in a different question and response format. We ask you to provide your thoughts on each. You may comment on any aspect, such as the phrasing or the way the item is scored or responded to.

Construct	Question	Response option(s)
1 Population alignment	A. The population from whom health utility estimates were elicited is well aligned with the population for which the economic evaluation is intended.	Yes No Unclear
	B. Respondents and economic evaluation population: *How well does the population from whom health utilities were elicited (i.e. respondents) fit with the context of the economic evaluation?*	Perfectly aligned Well aligned Some issues with alignment Incompatible Unable to evaluate Comments:
	Which format do you prefer? ☐A ☐ B ☐About the same ☐ Neither **Interviewee comments:**	
2 Year of study	A. Are there any concerns with respect to how recently the health utility study was conducted?	Yes No Unclear
	B. Year of study *Consider the difference from when the health utility study was conducted and when the economic evaluation is set.*	Sufficiently recent Issues with year of publication Unclear

Construct	Question	Response option(s)
	Which format do you prefer? ☐A ☐ B ☐About the same ☐ Neither **Interviewee comments:**	
3 Jurisdiction-specific guidance for health utilities	A. The methods used to elicit health utilities in the health utility studies are appropriate according to the jurisdictional preferences or restrictions.	Yes No Unclear
	B. Choice of health utility elicitation method *Some jurisdictions recommend or require specific health utility elicitation methods. If your jurisdiction has such a requirement or restriction, does the health utility study align?*	Yes No Unclear
	Which format do you prefer? ☐A ☐ B ☐About the same ☐ Neither **Interviewee comments:**	

Section 5: Adoption
Rationale: To gather participant opinions on how to best promote the HAT for adoption by researchers and decision-makers in real-world decision making.

5.1 Please share any feedback about the ways by which we may promote the HAT in real-world decision making.
[Comments]

Appendix B: Online survey questionnaire

The Health utility Application Tool – Development survey

Health utilities measure health-related quality of life or health status on a scale anchored at 0 for dead and 1 for full or perfect health. These utilities are used as quality weights in the calculation of the quality adjusted life years (QALYs). Health economic evaluations that use incremental cost per QALY gained, where the numerator is the incremental cost and the denominator the incremental QALYs between new and existing treatments, are known as cost utility analyses (CUAs). Owing to the generic nature of the QALY metric, the CUA allows for broad comparison across diseases, which is desirable in reimbursement decision making.

Reimbursement decision making is a population- and jurisdiction-specific process. Thus, health utilities used in CUA to inform decision making are ideally derived from the same context. However, published health utilities may be inappropriately used in a CUA (e.g. different patient populations or health states) or there may be poor compliance with decision maker's economic evaluation guidelines.

We are developing a novel tool, named the Health utility Application Tool (HAT), to promote the proper application of published health utilities in CUAs. We are seeking your input on issues that you have encountered with the use of health utilities. Your feedback will help us identify the most relevant and important issues that should be addressed with our tool.

Thank you very much for your consideration and participation.

Dr. Feng Xie
Dr. Gordon Guyatt
Dr. Holger Schünemann
Dr. Oren Levine
Dr. Yuan Zhang
Mr. Michael Zoratti
Department of Health Research Methods, Evidence, and Impact; McMaster University

Dr. Murray Krahn
Toronto Health Economics and Technology Assessment (THETA) Collaboration

Dr. Tammy Clifford
Canadian Institutes of Health Research (CIHR)

Mr. Donald Husereau
University of Ottawa

Dr. Kelvin Chan
Sunnybrook Health Sciences Centre

Demographics and professional experience

Gender
- Male
- Female
- Other
- Prefer not to say

Age range
- 18-30
- 31-40
- 41-50
- 51-64
- 65+

Education
- Bachelor's degree
- Master's degree
- Doctorate (Ph.D.)
- Medical degree (M.D.)
- Other degree(s) [please specify]

What year did you complete your most recent degree?

Who is your current employer?
- Academia
- Government/Non-profit organization
- Pharmaceutical industry
- Consulting
- Other

Which of the following best describes you?
- Health economist
- Non-health economist
- Other

How many years have you been in your current roe?
- <5 years
- 5-10 years
- 11-15 years
- More than 15 years

How many years have you been working in the general field of health economics and outcomes research?

- <5 years
- 5-10 years
- 11-15 years
- More than 15 years

Approximately how many drug reimbursement submissions do you encounter in a given year?

What is your role(s) in those encounters?

Issues with the use of published health utilities in cost-utility analyses

Here we present a list of issues related to the use of published health utilities in CUAs. Please rate their importance according to your own expertise and experience.

A systematic literature review was conducted to identify published health utilities.

1 – Not important at all
2
3
4
5 – Extremely important

The target patient population in the CUA is the same as from whom the health utilities were obtained.

1 – Not important at all
2
3
4
5 – Extremely important

The description of health states for which health utilities were measured reflects important characteristics of the health states considered in the CUA (e.g. economic model).

1 – Not important at all
2
3
4
5 – Extremely important

The methods used to measure health utilities (e.g. time trade-off, standard gamble, or utility-based
instruments such as the EQ-5D) meet the recommendations specified by the decision maker

1 – Not important at all
2
3
4
5 – Extremely important

The heath utility data was collected close to the year when the CUA was conducted.

1 – Not important at all
2
3
4
5 – Extremely important

If a utility-based instrument (e.g. EQ-5D or SF-6D) was used, a preference-based scoring algorithm
specific to the country that is the setting of the CUA was then used to calculate health utilities.

1 – Not important at all
2
3
4
5 – Extremely important

If multiple published health utilities were identified, the rationale of choosing from these

health utilities is provided.

1 – Not important at all
2
3
4
5 – Extremely important

If no health utility literature is available for the patient population studied in the CUA, a justification for using estimates for another clinical indication has been provided.

1 – Not important at all
2
3
4
5 – Extremely important

Please describe any other issues we have not addressed.

Sample questions

The way we ask questions is important.

Here we present a series of sample questions for three constructs that may be important in the final tool. Each set presents the same content (i.e. construct), though in a different question and response format.

For each, we ask that you provide your opinions on formatting (not the content) such as phrasing, the question's structure, and item response options.

Construct	Question	Response option(s)
Population alignment	A. The population from whom health utility estimates were elicited is well aligned with the population for which the economic evaluation is intended.	Yes No Unclear
	B. Respondents and economic evaluation population: How well does the population from whom health utilities were elicited (i.e. respondents) fit with the context of the economic evaluation?	Perfectly aligned Well aligned Some issues with alignment Incompatible Unable to evaluate Comments:

Based on this sample question, which format do you prefer?

- A
- B
- About the same
- Neither

Please comment on your preference.

Construct	Question	Response option(s)
Year of study	A. Are there any concerns with respect to how recently the health utility study was conducted?	Yes No Unclear
	B. Year of study Consider the difference from when the health utility study was conducted and when the economic evaluation is set.	Sufficiently recent Issues with year of publication Unclear

Based on this sample question, which format do you prefer?

- A
- B
- About the same
- Neither

Please comment on your preference.

Construct	Question	Response option(s)
Jurisdiction-specific guidance for health utilities	A. The methods used to elicit health utilities in the health utility studies are appropriate according to the jurisdictional preferences or restrictions.	Yes No Unclear
	B. Choice of health utility elicitation method Some jurisdictions recommend or require specific health utility elicitation methods. If your jurisdiction has such a requirement or restriction, does the health utility study align?	Yes No Unclear

Based on this sample question, which format do you prefer?

- A
- B
- About the same
- Neither

Please comment on your preference.

Adoption

Please share any feedback on the ways by which we may promote the HAT in real-world decision-making.

- Partnerships with HTA bodies
- Workshops at conferences
- Webinars
- Online tutorials
- Other, please specify

If you prefer to be acknowledged when we publish this tool, please provide your name and affiliation in the box below.

Name:
Affiliation
Email address:

CHAPTER 5 – CONCLUSIONS

This thesis explored the application of health utility inputs to health economic evaluations, with an emphasis on promoting practical, attainable solutions to methodological shortcomings. In culmination, these projects are intended to offer supports and guidance for investigators working in the field of HTA as they navigate difficult quantitative approximations of value.

The first proposed solution, the development of a registry of published health utility inputs, introduced as the HUB, is particularly intended to support decision makers with evaluating whether the evidence informing models reflects the literature. From the review of appraisal and reporting checklist tools, important gaps were identified in the formalized guidance for evaluating credibility and goodness of fit. Indeed, while one may grasp the principles and methodologies, there is little to support a formal, standardized approach to undertaking this task. This motivated the development of the HAT, a simple tool designed to promote transparency and communication in evidence selection. However, in order to remain relevant, these projects should not remain static. Rather, as the processes of HTA are dynamic, these projects too require a commitment to ongoing engagement with those in the field. Yet, they should also remain portable across applications and jurisdictions. As new evidence emerges, it should be identified and catalogued in the HUB registry. Similarly, new checklists and appraisal tools are likely to emerge and should be evaluated for their purpose and suitability to support both investigators and decision makers in their evaluations of the evidence. Finally, the HAT should be revisited regularly to ensure it appropriately reflects the needs of those in the

field. While the standard HAT model should remain indifferent to jurisdictional nuances, adaptations to suit the specific needs and interests of decision makers in specific contexts is highly encouraged.

The methods chosen to complete these thesis activities were selected to balance comprehensiveness and pragmatism. This is perhaps most significant for the development of the HAT, which was the product of three core research sub-projects. The first of these, as described in Chapter 3, was to evaluate the available literature to understand how previous work may be leveraged. From this, two engagement activities were planned. One of these was broadly scoped and invited researchers from any discipline, experience, or geography to participate in an online survey. The questions asked of respondents were framed based on the research team's understanding of the types of constructs we anticipated being reflected in the final tool, though with opportunities for open-ended responses to highlight additional issues and considerations. One limitation of this research activity was language, as the survey was only available in English. Despite this, several respondents from predominantly non-English countries participated and provided valuable feedback. The second peer engagement activity was the semi-structured phone interviews which were designed to directly engage with investigators who use or apply the results of economic evaluations in their daily work – individuals described as the doers and users of health utility evidence. Invitations for this activity were again broadly scoped to invite participants from academia, HTA bodies, and the pharmaceutical industry. While the response rate was lower than initially hoped from certain sectors, including the health technology assessment bodies, several methods experts participated

and offered valuable feedback which is directly reflected in the HAT as it has been described here.

Knowledge translation is key to the success of any research activity. This therefore became one of the pillars upon which these thesis activities were designed. For instance, the two engagement activities to support the development of the HAT were devised to reach a broad audience, not only as a means of soliciting feedback and opinions but also as an early phase of knowledge dissemination. Many participants, both in the online survey and interviews, agreed that the concepts we were presenting and discussing were worthwhile and that further methodological guidance was warranted. Many expressed an interest in supporting the development of a novel tool to achieve this. It is hoped that those participants critically engage with the HAT such that it can be reviewed and reframed over time.

At its core, this thesis is framed as a simple solution to a far-reaching methodological shortcoming. The impact of this work is yet to be realized and will require an ongoing commitment to continuous collaboration and active dissemination with peers. Promotion and active demonstration of these outputs will ensure their longevity and will shape them into relevant and impactful components of the health economist toolbox. The higher-level goals of the activities described in this thesis extend beyond the methods and ambitions of these papers. Expansion of the HUB registry beyond oncology is necessary. Similarly, the review of published tools highlighted addition limitations with respect to the critical appraisal of methodological rigour for these studies. It is hoped that this thesis becomes a launching point for further methods

investigations, thereby achieving its purpose of promoting accountability, transparency, and good practices in the field of HTA.