THE CONSISTENCY IN USING HEALTH UTILITIES BETWEEN COST-UTILITY ANALYSES IN ONCOLOGY AND REFERRED ORIGINAL HEALTH UTILITY STUDIES: A REGISTRY-BASED REVIEW

THE CONSISTENCY IN USING HEALTH UTILITIES BETWEEN COST-UTILITY ANALYSES IN ONCOLOGY AND REFERRED ORIGINAL HEALTH UTILITY STUDIES: A REGISTRY-BASED REVIEW

By ZHIYUAN CHEN, B.S., MA.Sc

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree Master of Science

MASTER OF SCIENCE (2019) McMaster University

(HEALTH RESEARCH METHODOLOGY) Hamilton, Ontario

TITLE:

The Consistency In Using Health Utilities Between Cost-Utility Analyses In Oncology And Referred Original Health Utility Studies: A Registry-Based Review

AUTHOR:

Zhiyuan Chen (McMaster University)

SUPERVISOR:

Professor Feng Xie

NUMBER OF PAGES: xi, 55

ABSTRACT

Background Cancer is a big threat to human health and imposes a heavy financial burden on health care systems worldwide. Cost-utility analyses (CUAs) have been widely used to measure cost-effectiveness of new cancer treatments. In a CUA, health utilities (HUs) are used to calculate quality-adjusted life years (QALYs) and thus play an important role in determining cost-effectiveness results.

Methods Oncology CUAs that were included in the Cost-Effectiveness Analysis (CEA) Registry of the Center for the Evaluation of Value and Risk in Health (CEVR), published in English through 2016 and reported HUs and QALYs were identified and included in the study. Data were collected from the oncology CUAs and referenced original HU studies. The consistency of HUs were assessed by comparing both health state descriptions and utility values between the oncology CUAs and corresponding original HU studies.

Results In total, 912 out of 1062 CUAs that investigated cancer diseases from the CEVR CEA Registry and 5583 HUs used in them were included in the analysis. 1353 HUs (24.2%) were measured along with the CUAs (defined as primary data) and 4230 HUs (75.8%) were derived from other sources (defined as secondary data). Out of the 3360 HUs for which the original studies were identified and compared with the CUAs, 1348 (40.1%) had the same health state descriptions and utility values, 633 (18.9%) differed only in value, 390 (11.6%) differed only in description, and 989 (29.4%) differed in both description and value. Among the 2012 HUs had either or both descriptions and utility values different from original HU studies, 377 (18.7%) of them were used without explanation, and 143 (7.1%) were derived from different diseases.

Conclusions Our study found there were discrepancies in using published health utilities in oncology CUAs. Consistent use of health utilities needs to be improved.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Prof. Feng Xie, who has spent enormous time and provided guidance on both the thesis project and my school life in McMaster. Without his persistent help, this project would not have been completed.

I would also like to show my gratitude to Zhou Ting from China Pharmaceutical University for his great help in this research; to Hannah McIntosh and other lovely staff in McMaster Health Sciences Library for patiently helping me gain access to dozens of papers for this large project; to Prof. Jim Bowen, Prof. Gord Blackhouse, and Prof. Oren Levine for their valuable comments on the thesis.

Last but not least, I would like to express my profound gratitude to my parents, Chen Shu and Zhang Qi, for their great love and unconditional support; thank my friends Sha Jizhou, Wang Mei, Jin Lin, Jin Xuejing, Yao Xiaomei, Jin Yanling and other schoolmates and teachers, in Canada and in China, who have always encouraged me and helped me throughout my years of study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CAD	Canadian Dollar
CEA	Cost-effectiveness Analysis
CI	Confidence Interval
CUA	Cost-utility Analysis
DSA	Deterministic Sensitivity Analysis
EQ-5D Questionnaire	EuroQoL Group 5-Dimension Questionnaire
HEE	Health Economic Evaluation
HRQoL	Health-related Quality of Life
HU	Health Utility
HUI	Health Utilities Index
ICER	Incremental Cost-effectiveness Ratio
ICUR	Incremental Cost-utility Ratio
LY	Life-year
NHIS	the National Health Interview Survey
NICE	National Institute for Health and Clinical Excellence
PSA	Probabilistic Sensitivity Analysis
Primary DID	Primary Disease ID
QALY	Quality-adjusted Life Year
QoL	Quality of Life
RS	Rating Scale
SA	Sensitivity Analysis
SD	Standard Deviation
SE	Standard Error
SG	Standard Gamble
CEVR CEA Registry	The Cost-Effectiveness Analysis Registry of the Center for the Evaluation of Value and Risk in Health
TTO	Time Trade Off
VAS	Visual Analogue Scale

DECLARATION OF ACADEMIC ACHIEVEMENT

I hereby declare that the research study presented in the form of my master's thesis was conceptualized and implemented by myself in full with academic inputs and suggestions from members of my thesis committee.

I also assume full responsibility for all omissions and errors that may have happened in spite of my scrutiny to the best of my efforts.

Zhiyuan Chen

CHAPTER I. INTRODUCTION

Cancer is the term given to a collection of related diseases characterized by the rapid creation, growth, and spread of abnormal cells in the body. There are more than 100 types of cancer, and each type is usually named for the organs or tissues where the cancer starts.[1] Cancer is one of the leading causes of morbidity and mortality worldwide[2] and, in 2018, there were an estimated 18.1 million new cancer cases and 9.6 million deaths globally[3, 4]. By 2030, there is expected to be 23.6 million new cancer cases per year.[2] According to Canadian Cancer Statistics, there were more than 200 thousand estimated new cancer cases and 80 thousand deaths due to cancer in Canada in 2017. Cancer is responsible for about 30.3% of all of deaths in Canada (based on 2012 estimates). About 1 in 2 Canadians are expected to be diagnosed with cancer in their lifetime. [5]

The rapid scientific progress in oncology has led to new diagnosis tools and therapies for cancer. These advancements have made a significant contribution to the early detection of cancers and improved survival rates. Although the number of newly diagnosed cancer cases have increased, the overall survival rates of cancer patients are increasing as well, which leads to the growing number of cancer survivors in Canada.[6] The age-specific lifetime cancer prevalence among Canadians aged 12 years and older increased slowly from 5.9% in 2005 to 6.8% in 2015.[6] However, apart from the advancements, new cancer therapies are more expensive. The total costs of cancer care in Canada increased steadily from $2.9 billion in 2005 to $7.5 billion in 2012 (in 2015 Canadian dollars (CAD)), with hospital care expenditures, physician care and drug expenditures accounting for the three largest portions [7]. In addition to these direct medical costs, indirect costs associated with cancer also result in a heavy financial burden on both patients and society, e.g. loss of income of both cancer patients and caregivers. A Canadian national wage loss associated with new cancer cases in 2009 was estimated to be $3.18 billion (in 2009 CAD).[8]

Cancer remains a big threat to public health and a heavy financial burden to both individuals and national health care systems. Advancement in cancer treatment is fast, and choices among alternative treatments in clinical practice are often needed. Therefore, it is requisite for all stakeholders, especially decision-makers, to be able to compare both the health and economic influences, or the cost-effectiveness, between multiple treatment strategies. As a result, health economic evaluation (HEE) has been routinely conducted to support reimbursement decisions for cancer-related interventions.

Cost-effectiveness evidence is required for a reimbursement decision to be made in many jurisdictions [9]. The health benefits in HEEs are usually measured using disease-specific outcomes or generic indicators such as life-years (LYs) gained or quality adjusted life years (QALYs). A HEE that uses QALYs as the outcome measure—thus incremental cost-utility ratio (ICUR) as the summary measure, is called a CUA. CUA is recommended by many decision makers worldwide due to its comparability across diseases.

In a CUA, the comparison between a new intervention and a standard care is made by a ratio of the difference in costs to the difference in QALYs:

$$ICUR = \frac{(C_1 - C_2)}{(QALY_1 - QALY_2)}$$

QALYs are calculated by weighting the time spent in a health state by the quality of life (QoL) weight associated with that state. HUs are cardinal preferences anchored at 0 (denotes death) and 1 (denotes perfect health) and are often used as QoL weights, which represent an individual's preferences for health states. HUs can be measured directly using established techniques or indirectly using pre-developed instruments [10, 11] The most commonly used direct preference measurement methods are the Standard Gamble (SG), time trade-off (TTO) and a variety of rating

scales (RSs). Pre-developed preference measurement instruments include the EuroQoL Group 5-Dimension（EQ-5D）Questionnaire, the Health Utilities Index (HUI), the Quality of Well-Being (QWB), and the SF-6D questionnaires.

The SG is a preference elicitation method based on the fundamental axioms of expected utility theory developed by von Neumann and Morgenstern in 1947 (also known as von Neumann-Morgenstern (vNM) utility theory).[12] The SG is considered a standard method for measuring health utilities with uncertainty.[13, 14] The TTO method was developed by Torrance et al. in 1972[15]. Both of the two methods are choice-based. The rating scales, on the other hand, request subjects to assign numerical values to provided health states.[13] Two commonly used rating scales are visual analogue scale (VAS) and category scaling. The rating scales are more straightforward and easier to use, compared to SG and TTO, especially in community-based samples. [13]

There are a number of preference-based instruments that are used to measure health utilities indirectly. One of the earliest preference-based instruments is the Quality of Well Being (QWB) developed in the 1970s.[16] The HUI and the EQ-5D, two most commonly used generic preference-based multi-attribute instruments, were developed in 1982 [17] and 1990 [18], respectively, and have been used increasingly since available. Generally, indirect methods are more feasible and time-efficient in measuring HUs than direct methods.

As one kind of the primary input data in CUAs, HU parameters are used to quantify the health benefits of interventions of interest, and thus could potentially impact the cost-effectiveness conclusions. Schackman et al. [19] reported that, after reviewing 94 sensitivity analyses (SAs) of HUs with quantitative results from 36 pharmaceutical CUAs, 41 (44%) HU SAs resulted in cost-utility ratios that were higher or lower than the specified thresholds compared with base-case results, suggesting CUA outcomes are sensitive to HUs. This is especially true when CUAs

investigate treatments for chronic conditions and long-term consequences.[20] As chronic diseases tend to impact patients' QoLs over a long period, both the time and HRQoL consist of the input parameters of health benefit measured in CUAs (i.e. difference in QALYs between different strategies), and thus, are influential to the cost-effectiveness outcomes.

As cancer treatments have effects on both patient's length and QoL, the application of HUs has a direct impact on the robustness of CUA outcomes, and is indirectly influential to health policy making. Most economic evaluation guidelines have an explicit, albeit general, statement about utility measurements [21], possibly due to a lack of standard methods. We are interested in finding out the application situation of HUs in published oncology CUAs.

We aimed to review the application of HUs in published oncology CUAs and assess the consistency in using HUs between the oncology CUAs and original HU studies. Chapter 2 describes a registry-based review of published oncology CUAs by descriptively analyzing their methodology characteristics, including perspective, time horizon and uncertainty analysis. Chapter 3 describes the characteristics of HUs used in the included CUAs and assessed the consistency between those CUAs and original studies, regarding both health state descriptions and utility values. The last chapter presents the limitations of the study and conclusions

CHAPTER II. REGISTRY-BASED REVIEW OF PUBLISHED COST-UTILITY ANALYSES IN ONCOLOGY

Previous guidance and reviews of health economic evaluations

It is important to ensure the quality of economic evaluations in health care, given their role in health policy making. There are a large number of guidelines providing recommendations for best practice of HEEs.[22~25] Edwards et al.[26] performed a systematic review of published guidance for HEEs between 1990 and 2012, in which 16 guidance papers were identified and reviewed. Given the variation in interventions (e.g. public health interventions versus clinical interventions) and the complexity of health care environment, a general consensus was reached that wider social and environmental costs and benefits should be taken into consideration, especially when it comes to public health interventions. Several key methodological challenges were mentioned, including the difficulty in evaluating costs and benefits of a specific public intervention in real world, the health inequity, and the insufficiency of using QALYs as the primary outcome of a complex intervention given related non-health outcomes and other relevant costs and effects.[26] Philips et al. investigated existing HEE guidelines extensively and built a general framework to help assessing the quality of HEEs.[22] The study reported disagreements among previous guidelines in three aspects: structure, data and consistency. Major contradictories among guidelines involved data identification (i.e. arguments about data derived from expert opinions in the study) and the assessment methods of parameter uncertainty.

Among all the diseases and therapies that need to be evaluated, cancer and its treatments is one of the challenging issues in health care. Since the 1990s, cost-effectiveness evidence of oncological treatments has been required to inform decision making for health care reimbursement in many jurisdictions, such as Canada and the UK.[27~29] The specific considerations when applying CUAs

in oncology have been investigated and discussed. A retrospective review by Yong et al.[30] evaluated the quality of economic evaluations of anticancer drugs reviewed by the Committee to Evaluate Drugs/Cancer Care Ontario (CED/ CCO) between April 2007 and March 2008, and concluded that 5 of the 15 included HEEs have methodology problems that prevent the advisory committee from using the analysis results. Those problems were classified into three categories: inappropriately estimating health benefits (e.g. estimating HUs or generalizing of clinical evidence), assuming costs in favor of manufactures, and lack of analysis validity (e.g. insufficient SAs). The Canadian Agency for Drugs and Technologies in Health (CADTH) provides specific guidance for the HEEs of oncology treatments.[27]

According to the previous studies by Edwards [26] and Philips [22], a consistent agreement regarding all the HEE methods recommended by the variety of guidelines has not been reached. Besides, there are still methodological challenges and issues found in guidelines and published HEEs, including the time horizon, complexity of decision models, data identification and assessment of uncertainty. [22, 23, 26,30~32] However, for cancer specifically, there is no review of methodology in published CUAs. The objective of this review was to identify and outline the main characteristics of published CUAs in oncology included in the CEVR CEA Registry [33].

Methods

Source of published CUAs in oncology

The CEVR CEA Registry (https://cevr.tuftsmedicalcenter.org/databases/cea-registry) is a database of more than 8,000 HEEs on various diseases and treatments published since 1976.[33] The CEVR CEA Registry searches all HEEs indexed in MEDLINE with keywords including "QALYs", "quality" and "cost-utility analysis". The major inclusion criteria used by the registry are original

HEEs published in English and measuring health benefits using QALYs. After abstract screening, eligible studies were included for full-text review, information of interest will be collected by 2 trained reviewers independently on 3 aspects: 1) article and methods; 2) ratio and 3) utility weight.[33]

According to the registry, the classification of diseases is determined by a clinician and labeled by the variable "Primary Disease ID" (Primary DID). Among all those categories, the diseases of interest consist of malignant neoplasm, colorectal cancer, lung cancer, breast cancer, cervical cancer, uterine cancer, ovary cancer, prostate cancer, hematologic cancers (lymphomas and leukemia) and other neoplasms.[33] Our study included all CUAs that were classified as cancer-related diseases in the primary disease categories in the registry.

For the purpose of this study, all the information from the Article and Methods Section and the Utility Weight Section from the registry was downloaded for further use via sponsorship access (Academic and Non-profit Partners). Data from the registry was downloaded, filtered (by the Primary DID) and organized in the STATA 14 software[34] (details in Appendices Table S1).

Study inclusion and data collection

All the CUAs in cancer identified in the CEVR CEA Registry were reviewed in full text. The inclusion criteria were being included in the CEVR CEA Registry, being original CUAs published between 1976 and 2016, written in English, investigated cancer or neoplasm as their primary diseases, and reported at least one health state description and utility. No other limit was performed.

We defined original HU studies as those which measured and reported HUs. A CUA might contain original HU studies or obtain HUs from external sources. In the latter case, we identified and

retrieved the original health studies from external sources whenever possible to compare the HUs used in the CUA with those reported in the original studies.

The final data extraction contents consisted of the basic study information, demographics of the patient population, the treatment information, the information related to the CUA and HUs (Table 1).

The basic study information includes publication year, perspective, time horizon, and the target cancer classifications according to the CEVR CEA Registry. We also extracted additional study characteristics including the study design, the treatment stage, the number of HUs and uncertainty analyses from the CUA. For the treatment stage of each study, it was classified into six categories: 1) prevention and diagnosis stage, including prevention, screening, gene mutation tests, surveillance and diagnosis for new cancer cases, and staging or assessing the node/metastasis status in cancer patients; 2) first-line treatment; 3) second- and/or third-line treatment, including best supportive care; 4) adjuvant/neoadjuvant treatment for cancer recurrence; 5) post-treatment follow-up, e.g. watchful waiting and rehabilitation; and 6) other treatment stages, including unstated stage and multiple treatment lines. The classification was judged by searching for the related statements, e.g. “first line”, “newly diagnosed”, “adjuvant”, “best supportive care”. If the key phrases were not available, the judgment would be made by the author depending on the treatment goal, target population and interventions in CUAs. For the number of HU references per study, all the studies that were referenced in the oncology CUAs (with a reference number) for HU inputs, whether or not reported HU indexes directly, were counted. For the uncertainty analysis, whether or not the deterministic and probabilistic SAs (DSAs and PSAs) were performed were found out, especially for HU parameters. If there was no DAS conducted in a CUA, then whether or not a case-scenario analysis involving HU parameters were reviewed.

The review results were presented in three continuous time periods (1988-2000, 2001-2010 and 2011-2016) and then in total numbers, respectively. As there were limited oncology CUAs identified before 1990, those studies were combined to the next decade and analyzed as the first time period (1988-2000). As there were different numbers of studies in each time period, the proportion and number of each method characteristic were both used in results. In this format, the change trend of characteristics in oncology CUAs over the last almost three decades (from 1988 to 2016) could be compared between each time period.

Data extraction, management and descriptive analyses were performed using a pre-developed and tested data extraction form in Microsoft Excel, Version 16.16.9 (2018).

Table 1. Data extraction form of CUAs

Study information and characteristics	Data usage
Title, abstract, first author, publication year, study countries and regions [i], sponsorships, target cancer classification, perspective, time horizon and ICUR.	Provided by the CEVR CEA Registry and used in analysis directly
Study design: if any decision analytic model was used, e.g. the Markov model. The target population. Treatment stage of the target intervention [ii]: the line of treatment. Number of HUs (with utility values reported) Total number of references of HUs (assigned a reference number in text). [iii] Uncertainty analysis on HUs [iv]	Additional data extracted

[i] The study countries and regions refer to the geographical jurisdiction to which the study's results were applied.
[ii] The treatment setting was classified into the following types: prevention/diagnosis, first line; second/third line, adjuvant/neoadjuvant treatment; post-treatment follow-up and others.
[iii] The references of HUs refer to all the references related to the HUs that are referenced in the CUAs.
[iv] Uncertainty analyses include deterministic and/or probabilistic sensitivity analyses and scenario analyses.

Results

As shown in Figure 1, there were a total of 1062 articles published between 1976 and 2016 included in the registry under the primary disease of cancer or neoplasm (the database was assessed on May 1, 2018). One study was removed as duplicate. Then 149 CUAs were excluded after screening with 122 reporting no HU value, 24 that are not CUAs or original CUAs, 1 review of economic evaluations, 1 study not in English and another one investigating non-cancer disease. As a result, 912 CUAs (85.9%) were included in the review, all assessed in full text with assistance from the Health Science Librarians at McMaster University (details in Appendices Table S2).

The main characteristics of the included CUAs were summarized in Tables 2-4.

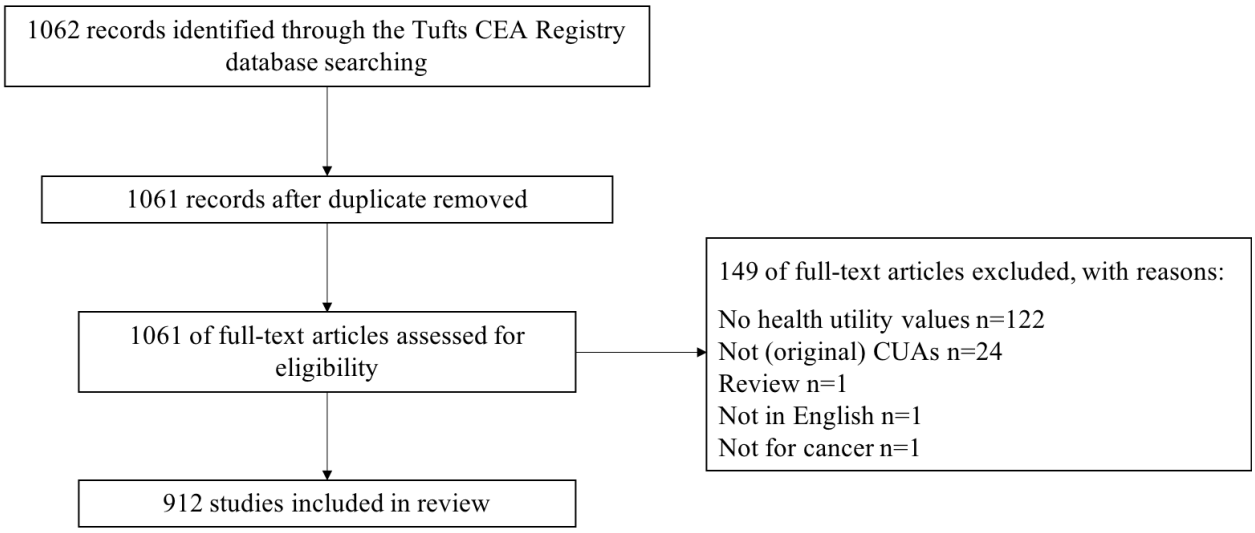


Figure 1. Flow chart of the registry-based review

Characteristics of the CUAs in oncology

Publication year Although the registry database aims to include CUAs that were published since as early as 1976, the first CUA in oncology was not identified until 1988. In general, the number of published CUAs in oncology has increased over time, especially since 2005 (Figure 2).

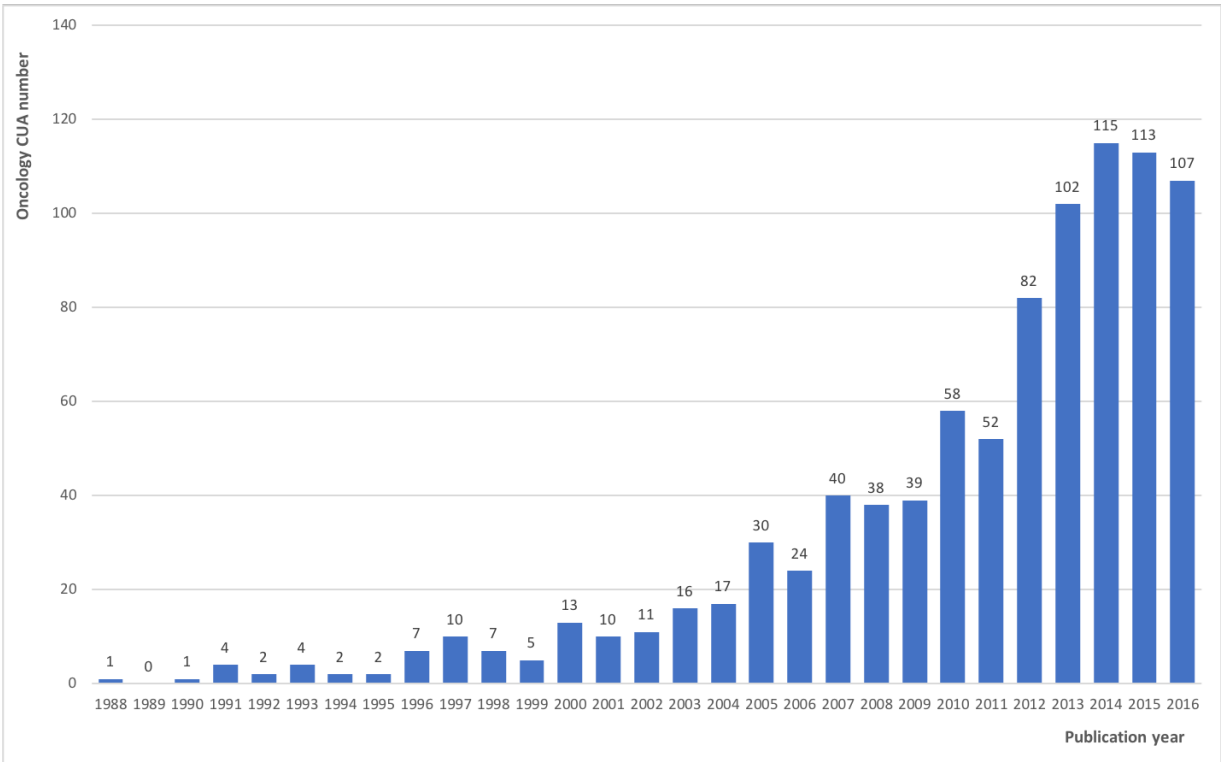


Figure 2. Number of published oncology CUAs by year

Study country and sponsorship The numbers and proportions of study countries and sponsorships are listed in Table 2. About 40% (366 out of 912) of published CUAs for cancer were performed in the United States (USA). The two most common sponsorship types were government organizations and industry (pharmaceutical and medical device companies), each accounting for about 27% of all the funding sources.

Table 2. Study country and sponsorship of the 912 CUAs in oncology [i]

Countries and regions	Number	Proportion	Sponsorships	Number	Proportion
US	366	40.1%	Government	249	27.3%
Canada	103	11.3%	Industry [iii]	248	27.2%
UK	99	10.9%	No information	186	20.4%
Netherlands	53	5.8%	Foundation	89	9.8%
China [ii]	40	4.4%	None	78	8.6%
Australia	24	2.6%	Health care organization	39	4.3%
France	24	2.6%			
Italy	21	2.3%	Professional membership organization	20	2.2%
Sweden	20	2.2%			
Others	208	22.89%	Others	75	8.2%
Total	949	104.1%	Total	984	107.9%

[i] One CUA could have more than one study country and/or one type of sponsorship. Thus, the total number of study countries and sponsorships are both more than 912.
[ii] Excluded Taiwan.
[iii] The industry refers to pharmaceutical and medical device companies.

Target cancer Table 3 shows that the most frequently investigated type of cancer in the published oncology CUAs was malignant neoplasms (69.3%), followed by breast cancer (13.6%).

Table 3. Primary cancer types of the 912 CUAs in oncology*

Primary Disease ID in the registry	Cancer category	Number	Proportion
2	Malignant neoplasms	632	69.3%
122	Breast cancer	124	13.6%
127	Hematologic cancers (Lymphomas, Leukemia)	35	3.8%
128	Other neoplasms	30	3.3%
121	Lung cancer	28	3.1%
120	Colorectal cancer	27	3.0%
126	Prostate cancer	23	2.5%
125	Ovary cancer	8	0.9%
123	Cervical cancer	5	0.5%

* The primary cancer types were classified by the CVER CEA Registry without detailed definition.

Model of the 912 CUAs 840 CUAs (92% of 912) used model. 539 model-based CUAs (64% of 840) used Markov model, making this the most commonly used model type. A decision tree was often used in combination with the Markov model (60 out of 840, 7.1%) in these studies (Table 4).

Table 4 Characteristics of included CUAs in oncology published between 1988 and 2016

Basic characteristics	1988-2000 (n=58)	2001-2010 (n=283)	2011-2016 (n=571)	Total (n=912)
Study type				
Model-based study	53 (91.4%)	267 (94.3%)	520 (91.1%)	840 (92.1%)
Trial-based study	5 (8.6%)	16 (5.7%)	51 (8.9%)	72 (7.9%)
Perspective [i]				
Health care payer	43 (74.1%)	229 (80.9%)	429 (75.1%)	701 (76.9%)
Societal	15 (25.9%)	51 (18.0%)	15 (2.6%)	81 (8.9%)
Others [ii]	0 (0)	0 (0)	66 (11.6%)	66 (7.2%)
Not stated / Unclear [iii]	0 (0)	3 (1.1%)	61 (0.5%)	64 (0.7%)
Time horizon [i]				
Lifetime	28 (48.3%)	145 (51.2%)	253 (44.3%)	426 (46.7%)
Non-lifetime	17 (29.3%)	122 (41.7%)	251 (44.0%)	390 (42.8%)
Not stated	13 (22.4%)	16 (7.1%)	67 (11.7%)	96 (10.5%)
Intervention setting				
Prevention/diagnosis	12 (20.7%)	46 (16.3%)	130 (22.8%)	188 (20.6%)
First-line	7 (12.1%)	41 (14.5%)	122 (21.4%)	170 (18.6%)
Second and/or third line(s)	7 (12.1%)	47 (16.6%)	68 (11.9%)	122 (13.4%)
Adjuvant/neoadjuvant treatment	12 (20.7%)	68 (24.0%)	64 (11.2%)	144 (15.8%)
Post-treatment follow-up	1 (1.7%)	10 (3.5%)	17 (3.0%)	28 (3.1%)
Others	19 (32.8%)	71 (25.1%)	170 (29.8%)	260 (28.5%)
Median (range) of HU number in each CUA	4 (1-24)	4 (1-27)	4 (1-46)	4 (1-46)
Median (range) of HU references number in each CUA	1 (0-5)	2(0-18)	2 (1-77)	2 (0-77)
ICUR				
Reported	55 (94.8%)	276 (97.5%)	488 (85.5%)	819 (89.8%)
Not reported	3 (5.2%)	7 (2.5%)	25 (4.4%)	35 (3.8%)
Unclear [iii]	0	0	58 (10.2%)	58 (6.4%)
DSA/Scenario analysis				
DSA incorporated HUs	25 (43.1%)	189 (66.8%)	391 (68.5%)	605 (66.3%)
DSA did not incorporate HU	10 (17.2%)	53 (18.7%)	76 (13.3%)	139 (15.2%)
No DSA, but scenario SA tested HUs	16 (27.6%)	24 (8.5%)	52 (9.1%)	92 (10.1%)
Others	7 (12.1%)	17 (6.0%)	52 (9.1%)	76 (8.3%)
PSA				
PSA incorporated HUs	0 (0)	133 (47.0%)	383 (67.1%)	516 (56.6%)
PSA did not incorporate HUs	0 (0)	14 (4.9%)	13 (2.3%)	27 (3.0%)
No PSA	58 (100%)	136 (48.1%)	175 (30.6%)	369 (40.5%)
HUs were not tested in neither DAS nor PSA	33 (56.9%)	56 (19.8%)	94 (16.5%)	183 (20.1%)

HU Health utility; ICUR Incremental cost-utility ratio; DSA Deterministic sensitivity analysis; PSA Probabilistic sensitivity analysis; SA sensitivity analysis.

[i] Information of study perspective and time horizon are provided by the CEVR CEA Registry.

[ii] "Other perspective" includes limited societal, health care sector and other perspectives.

[iii] "Unclear" resulted from missing data from the CEVR CEA Registry.

Perspective The health care payer perspective was used in about 80% (701 out of 912) of CUAs, making it the most frequently used perspective. The proportion of cancer CUAs that used the societal perspective decreased continuously from 25.9% (1988-2000) to 18.0% (2001-2010) and finally to 2.6% (2011-2016) (Table 4).

Time horizon The proportion of oncology CUAs that adopted life time horizon was relatively stable over the three time periods – ranging between 44.3% and 51.2% (Table 4).

Treatment stage Our results show that, except for other treatment stages (n=260, 28.5%), cancer prevention and diagnose stage, was the most commonly studied stage, accounting for 29.5% (269 out of 912) of oncology CUAs (Table 4).

HU The total number of HUs extracted from the included CUAs were 5583, excluding anchor HUs (i.e. HUs equal to 1 or 0, denotes perfect health or death, respectively). From 1988 to 2016, the maximum number of HU parameters in the CUAs increased from 24 to 46, while the median number remained the same, which was 4 (Table 4). The number of HU references in one CUA (recorded in reference lists) increased from 5 in 1988 to 77 in 2016. The median number of HU references was 1 in 1988-2000 and 2 in 2001-2010 and 2011-2016.

ICUR Only about 35 (3.8% of 912) of the oncology CUAs did not report ICUR in their results. Furthermore, according to the CEVR CEA Registry, only 2 of the 35 studies did not report sufficient data for readers to calculate the ICUR.

Sensitivity analyses As shown in Table 4, overall, out of 912 CUAs, 744 (81.6%) and 543 (59.5%) of them performed deterministic SAs (DSAs) and PSAs, respectively. The proportion of oncology CUAs performing DSAs increased from 60.3% (1988-2000) to 85.5% (2001-2010) and to 81.8% (2011-2016). The proportion of those performing PSAs increased largely from 0 (1988-2000) to 52.0% (2001-2010) and to 69.4% (2011-2016) over the time periods. 183 out of 912 (20.1%) CUAs did not incorporated HUs in either DSA or PSA. This proportion deceased from 56.9% (1988-2000) to 19.8% (2001-2010) and to 16.5% (2011-2016).

The proportion of oncology CUAs conducted DSAs increased from 43.1% (1988-2000) to 66.8% (2001-2010) and to 68.5% (391 out of 571). The proportion of those performing PSAs increased from 0 (1988-2000) to 47.0% (2001-2010) and 67.1% (2011-2016). The proportion of oncology CUAs performed scenario SAs for HUs that did not perform DSA incorporating HUs decreased from 27.6% (1988-2000) to 8.5% (2001-2010) and slightly increased to 9.1% (2011-2016). In total, there were 10.1% (92 out of 912) of this kind of CUAs.

Discussion

This chapter reviewed the basic information and methodology of 912 oncology CUAs published between 1976 and 2016 and included in the CEVR CEA Registry database. Overall, the number of published CUAs in oncology has been increasing significantly since 2005 – the publication number of oncology CUAs in 2005 was twice more than the average publication number between 2001 and 2004, and 6 times more than that in the first time period (1988-2000). The US, Canada and the UK are the top three countries where these CUAs were conducted. The sponsorships from government and medical industry were the two most common funding sources. Among all the cancer diseases, malignant neoplasms and breast cancer were the top two types of cancer that had

been investigated as the primary disease. The majority of these CUAs were model-based, adopted the health care payer's perspective, and reported ICURs.

The growing trend of the quantity of oncology CUAs (Figure 2) in this study is consistent with that of general CUAs overall from previous studies[22,35]. The number of HU parameters and the number of related references in published oncology CUAs had both increased largely over time as well.

According to the primary disease classification by the CEVR CEA Registry, apart from malignant neoplasms, breast cancer has received the most attention. When it comes to the treatment setting, the increasing proportion of CUAs on cancer preventions or diagnoses reflects the fact that the early detection and prevention technologies of cancer have received growing attention. This is likely due to the development of new diagnostic and preventative technologies in recent years, such as the gene mutation detect assays and cancer prevention vaccines.

Most guidelines recommend a life-time horizon for HEEs in order to capture all the important differences between options.[22] The CADTH specific guideline for oncology products recommends using a lifetime horizon by extrapolating data using acceptable modelling techniques.[27] Despite the fact that cancer has a long-term effect on patient lives, less than half of those published CUAs adopted a lifetime time horizon. This is probably related to the poor prognosis of some cancer patients and the lack of long-term follow-up data. Additionally, the CEVR CEA Registry does not indicate the detailed method that was used to classify the CUAs as adopting the life-time horizon or not, as a result, the proportion of oncology CUAs did not use a lifetime horizon, could be overestimated. For example, the proportion of the 5-year relative survival rates for people diagnosed with distant-stage cancer is usually less than 30%.[36, 37] It is possible that CUAs focusing

on such cancer patients with poor prognosis used a time horizon of only several decades and were not recorded as a "lifetime" horizon in the registry.

The results of a CUA could be largely affected by uncertainty if not handled properly. The parameter uncertainty of individual HU parameters used in CUAs in oncology can be assessed by DSAs, while the overall parameter uncertainty and decision uncertainty can be assessed by PASs.[24] Based on our results, the proportion of oncology CUAs performing uncertainty analyses of HU input has not been high, either, with only 66.3% (605 of 912) of which performed DSA and 56.6% (516 of 912) performed PSA. Even in the latest period (2010-2016), the proportions of CUA incorporating HUs in DSA and PSA were both less than 70%. The study by Yong JH et al. demonstrated that, from the reviewers' perspective, one of the common issues was that insufficient sensitivity analysis would weaken readers' confidence in economic outcomes and hinder the cost-effective results from being used.[30] Our results confirmed this problem.

Additionally, our findings indicate that the ranges of HUs used in DSAs were sometimes arbitrary. For example, the CUA of lung carcinoma by de Lima Lopes et al.[38] varied all the input parameters, including HUs, in the range of 50% to 200%, without specifying any justification for using that range. In another case, the CUA of breast cancer by Pandharipande et al.[39] tested the all the 4 utility values used in DSA. The 4 health states are systemic chemotherapy, tamoxifen treatment, metastatic breast cancer and post-adjuvant therapy, with the utility value (and SA ranges) equal to 0.72 (0.5-1), 0.82 (0.5-1), 0.4 (0.2-1) and 0.92 (0.5-1), respectively. The SA ranges of the 4 HUs are different from their mathematic dispersions (e.g. SD, 95% CI or interquartile) in original studies when available. In this condition, the SA results could be confusing when being translated.

Our findings also indicate, oncology CUAs that target treatments for certain cancer types were relatively infrequently studied, including cervical cancer, ovary cancer, prostate cancer and uterine

cancer. This type of CUAs are expected and probably needed in the future. When conducting a CUA for cancer treatments, health economists should use a lifetime horizon, or use a time horizon long enough to capture all the possible life expectancy with reasonable explanation, and perform both DSA and PSA for all key input parameters, as recommended by guidance. Reporting transparency and necessary explanation (e.g. for varied ranges in SAs) are always needed for the purpose of application. Future studies and guidance of determining the appropriateness of varied ranges for uncertainty analysis in CUAs are welcomed.

Conclusion

We found there is an increasing interest in conducting CUAs in oncology over time. The majority of oncology CUAs used the model-based approach, adopted the health care payer's perspective, and reported ICURs but lacked the use of long-time horizon and sufficient SA related to HUs.

CHAPTER III. CONSISTENCY IN USING HEALTH UTILITIES BETWEEN PUBLISHED CUAS IN ONCOLOGY AND REFERRED ORIGINAL HEALTH UTILITY STUDIES

Previous guidance and reviews of health economic evaluations

HU is a key parameter for estimating the health benefits and cost-effectiveness outcomes in CUAs. HUs can be measured directly using established techniques (e.g. the SG and TTO),[12, 15] or indirectly using pre-developed instruments (e.g. the EQ-5D and the HUI).[17, 18] Although multiple guidelines recommend standard methods of implementing a CUA[21, 40], there are still issues and challenges found during the procedure of using HUs. Previous reviews investigated various aspects in the application of HUs and production of utility-weighted QALYs.[41~43] Arnold D et al. reviewed and compared health states measured by both the direct and indirect methods (SG and TTO vs EQ-5D, SF-6D and HUI) from same patients in previous studies. Jacek et al. reviewed 3 indirect measurement methods (QWB, EQ-5D and SF-6D) by comparing hypothetical health states head-to-head. Their findings indicate that utility values measured by different methods could be substantially different from each other [42,]. Roberta et al. suggested that the HUs used in one CUA should be obtained from the same data source or measured by the same method, and the final choice for the data source and any other adjustments made should be explained clearly.[41] Other methodological issues in published CUAs found in previous large reviews include the lack of transparency and consistency in reporting, especially in those published before 2001[35, 44]; and insufficient SA of inputs [30]. We are interested in looking for issues in the application of HUs in published CUAs in oncology, from early to recent years, and comparing the methodology practice over time to capture any changes.

According to the systematic reviews of published guidance for HEEs, by Edwards[26] and Philips[22], a consistent agreement regarding all the HEE methods recommended by the variety of guidelines has not been reached. There are still methodological challenges and issues found in guidance and published HEEs.[30] For example, adoption of time horizon, complexity of decision models, data identification and assessment of uncertainty. [22, 23, 26, 30, 45, 46] When it comes to HUs, generally, referencing HU sources, justifying the derivation methods for utility values when needed, and testing the parameter uncertainty all contribute to the "good practice" of HEEs. While at the same time, the standard approach of justifying the derivation methods for HU weights, deriving HUs from expert opinions and conducting SAs regarding HUs (e.g. determining the varied ranges of HU weights) remain unclear and need further investigation. [22]

The objective of this chapter is to review the application of HUs in the published CUAs in oncology, by descriptively analyzing the data sources, references, measurement methods and respondents of those HUs, and assessing the consistency of health state descriptions and utility values between those used in CUAs and those from the original HU studies.

Methods

Inclusion criteria and identification of original HU studies

All the HUs used in the 912 oncology CUAs included (see Chapter 2), except anchor states of full health (equal to 1) and death (equal to 0), were extracted if they were reported with specified utility values.

We defined the secondary HUs as those were first measured in previous studies and then used in oncology CUAs; and the primary HUs as those measured and used along with the CUAs. We

identified the original HU studies of these secondary HUs by tracking corresponding references in the oncology CUAs.

During the tracking process, the full text of referenced studies and all other studies went through were accessed, identified original studies were downloaded in full-text for further data collection, whenever available. Ideally, the original HU studies would be retrieved by simply accessing the referenced studies in oncology CUAs. However, if the referenced studies used secondary HUs, too (e.g. another CUA), then their references would be tracked again until the original HU studies were obtained, or not accessible. For each HU, the process of locating its original HU study was recorded by a number named "reference level", which was defined as the number of referenced studies went through to locate its original study.

In other words, if a secondary HU was referenced correctly, its reference level would equal to 1. Otherwise, its reference level would be 2 or more. The "reference level" can indicate the directness and accuracy of HU references in oncology CUAs.

Data extraction

The data extraction was performed by one reviewer (ZC) in a pre-developed and tested template. The form collected information on descriptions of health states, utility values, statistical ranges (e.g. 95% confidence interval (CI), standard error (SE), or standard deviation (SD)), measurement methods, the demographics of respondents, reference levels, publication years and first authors of original HU studies (last 3 items applied to secondary HUs only) (Table 5).

Corresponding information from the CEVR CEA Registry was used as comparisons to ensure the validity of data extraction, including but not limited to health states, utility values, statistical ranges, measurement methods, sample size and type of respondents (Table 5).

Table 5. Data Extraction Form of HUs

HU	**Extracted data from CUAs**	**Extracted data from original HU studies**
Primary HU: measured in oncology CUA	Health state description, utility value, varied ranges tested in SAs, measurement method(s), the demographics of respondents	N/A
Second HU: measured in and obtained from external source(s)	Study information (collected before in Chapter 2)	First author and publication year of the original HU studies, reference level, target disease; Corresponding health state description, utility value, statistical ranges (e.g. 95% CI, SE, or SD) measurement method(s), the demographics of respondents
	Health state description, utility value, varied ranges tested in SAs	

HU health utility; 95% CI 95% confidence interval; SE standard error; SD standard deviation

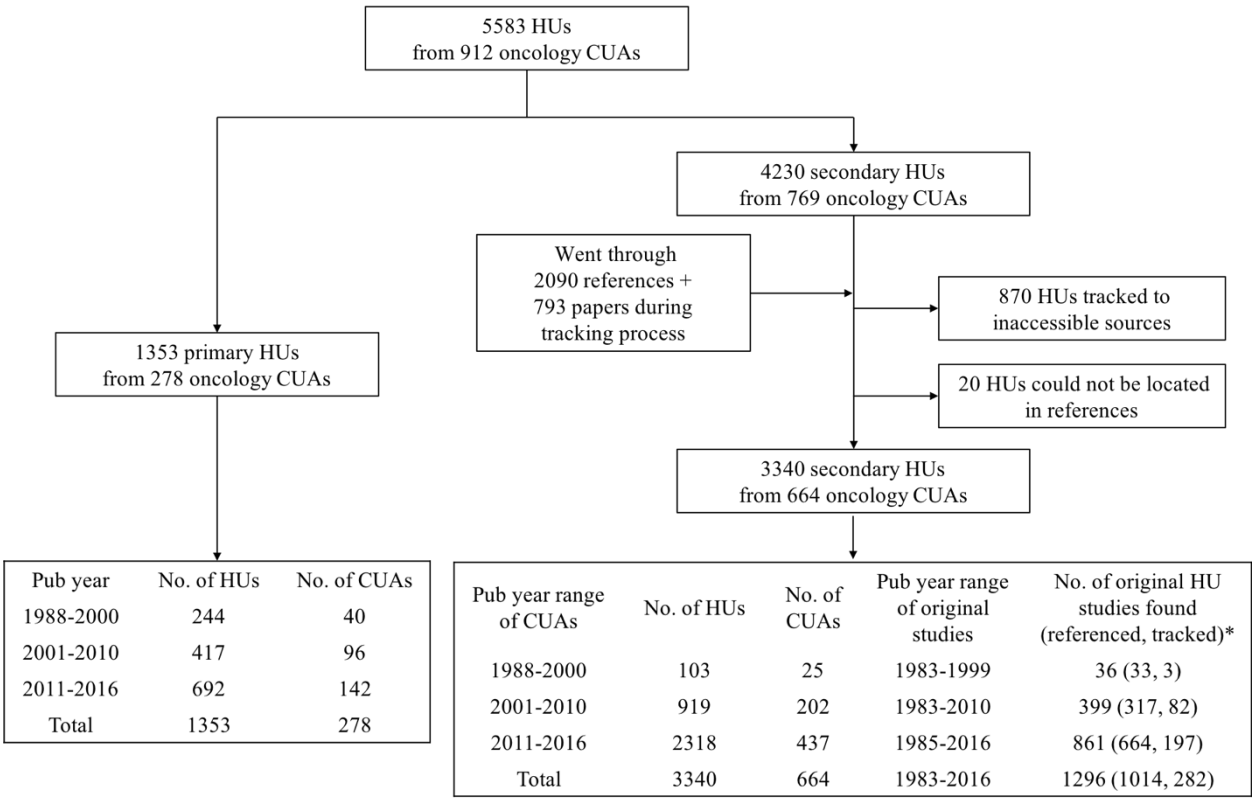
Consistency of secondary HUs

The HU consistency was assessed by comparing both the HU descriptions and utility values of each HU parameter between the oncology CUAs and original HU studies, given adjustments and assumptions on input parameters are often needed.

All the data collection, management and consistency assessments were performed in Microsoft Excel, Version 16.16.9 (2018).

Results

In total, 5583 HUs with specified utility values were extracted from the 912 published CUAs in oncology identified in the CEVR CEA Registry for descriptive analysis. The diagram of the review process was shown in Figure 3. The main characteristics of the collected HUs were shown in Tables 6-7 and Figures 4-7.



Pub year	No. of HUs	No. of CUAs
1988-2000	244	40
2001-2010	417	96
2011-2016	692	142
Total	1353	278

Pub year range of CUAs	No. of HUs	No. of CUAs	Pub year range of original studies	No. of original HU studies found (referenced, tracked)*
1988-2000	103	25	1983-1999	36 (33, 3)
2001-2010	919	202	1983-2010	399 (317, 82)
2011-2016	2318	437	1985-2016	861 (664, 197)
Total	3340	664	1983-2016	1296 (1014, 282)

Figure 3. Diagram of the consistency assessment of HUs used in oncology CUAs

* "Referenced" HU studies refer to the original ones that were referenced directly in CUAs (reference level=1).
"Tracked" HU studies refer to those being identified during the tracking process, with reference levels larger than 1.

Overview of reviewing process

As shown in Figure 3, 1353 of the 5583 (24.2%) HUs were measured alongside 278 CUAs, 4230 (75.8%) came from published literature in 769 CUAs (the median number of references was 4). During the process of identifying original studies of HUs, there were totally 2883 references went

through (2090 referenced studies in CUAs and 793 references tracked through reference lists whenever available). Furthermore, the number of corresponding original HU studies that were finally tracked to was narrowed down to 1296, in which the original HU data was measured and reported. As one HU original study could be referenced by multiple CUAs, there could be overlaps between and in each time interval.

Characteristics of the HUs used in oncology CUAs

The proportion of oncology CUAs that used primary HUs decreased from 69.0% (1988-2000) to 33.9% (2001-2010) and to 24.9% (2011-2016). On the contrary, the proportion of oncology CUAs that used secondary HUs, which increased from 43.1% (1988-2000) to 71.4% (2001-2010) and to 76.5% (2011-2016) (Figure 3 and Table 6).

Publication year For the secondary HUs, the earliest publication years of the original HU studies that were tracked to were 1983 (Figure 3).

Original sources and reference levels 212 out of 5583 (3.8%) of all the extracted HUs were cited with incomplete or no reference. These 212 HUs were extracted from 40 different oncology CUAs (4.4% of 912). 462 (8.3%) were tracked to inaccessible sources, including unpublished sources, abstracts, books, posters, reports not in English, and so on. These 674 secondary HUs, combined with 196 (3.5%) ones tracked to studies without HU data (totally n=870, 15.6%), were not analyzed afterwards due to lack of information.

For the reference level in Table 6, our analyses only included HUs that could be tracked to reviews or individual studies, the numbers of which were summed to 3340. The maximum reference level tracked was 4. The proportion of 1st-level reference is 78.0% (2606 out of 3340), while that of 2nd or more level is 22.0% (734 out of 3340).

Measurement methods To assess the measurement methods of the extracted HUs from oncology CUAs, all the primary HUs and secondary ones from accessible published reviews and individual studies were included. The total number was summed to 4693. As shown in Table 6, the subgroup classifications consisted of direct and indirect measurement methods, other methods (expert opinions and author assumptions), multiple methods and unclear ones. 2303 of the 4693 HUs (49.1%) were measured directly, making it the largest method classification. The proportion of HUs measured directly reduced from 56.2% (1988-2000) to 54.8% (2001-2010) and to 45.7% (2011-2016), while that measured indirectly increased fast from 7.8% (1988-2000) to 25.4% (2001-2010) and to 38.9% (2011-2016). The proportion of HUs determined by expert opinions and author assumptions declined from 36.0% (1988-2000) to 18.1% (2001-2010) and to 13.5% (2011-2016).

As shown in Figure 4, for each kind of specified measurement method, the most frequently used direct and indirect methods were the SG and EQ-5D, taking up 27.7% (n=1298) and 19.8% (n=928) of all the 4693 HUs, respectively. Apart from the commonly used methods, "Other direct methods" (n=47, 1.0%) consisted of the Disability Weight (n=18), Reference Gamble (n=7), Direct Question Objective (DQO, n=1), and single question asking participants to assign HU values (n=21). "Other indirect methods" (n=39, 0.8%) consisted of the Quality of Well Being (QWB, n=29), Assessment

of Quality of Life (AQoL, n=4), 15D (n=3), Utility Based Quality of Life Questionnaire-Cancer (UBQ-C, n=2) and Quality of Life in Reflux and Dyspepsia Patients (n=1).

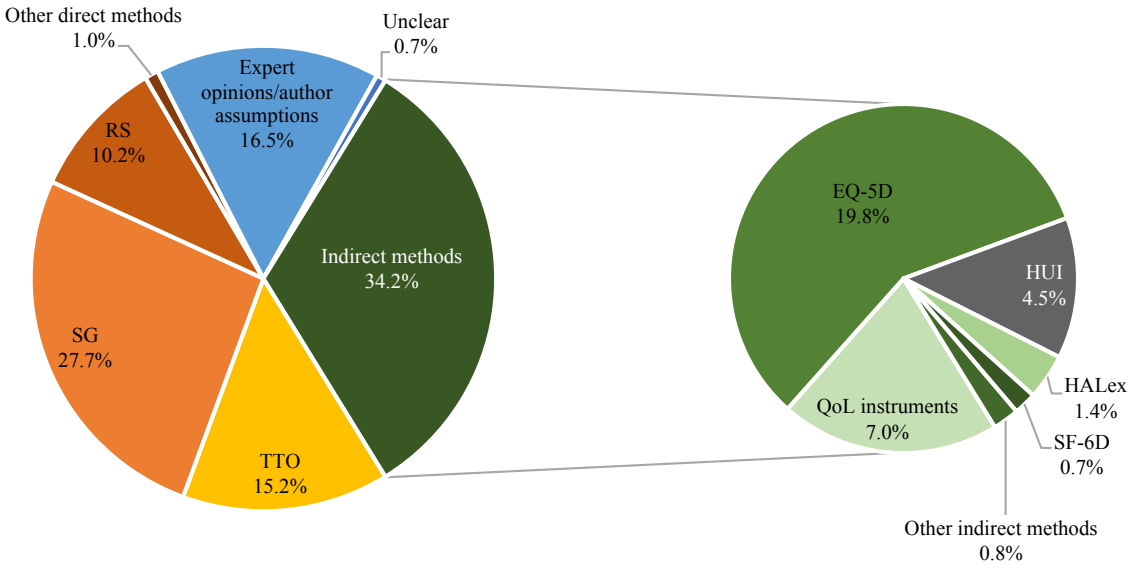


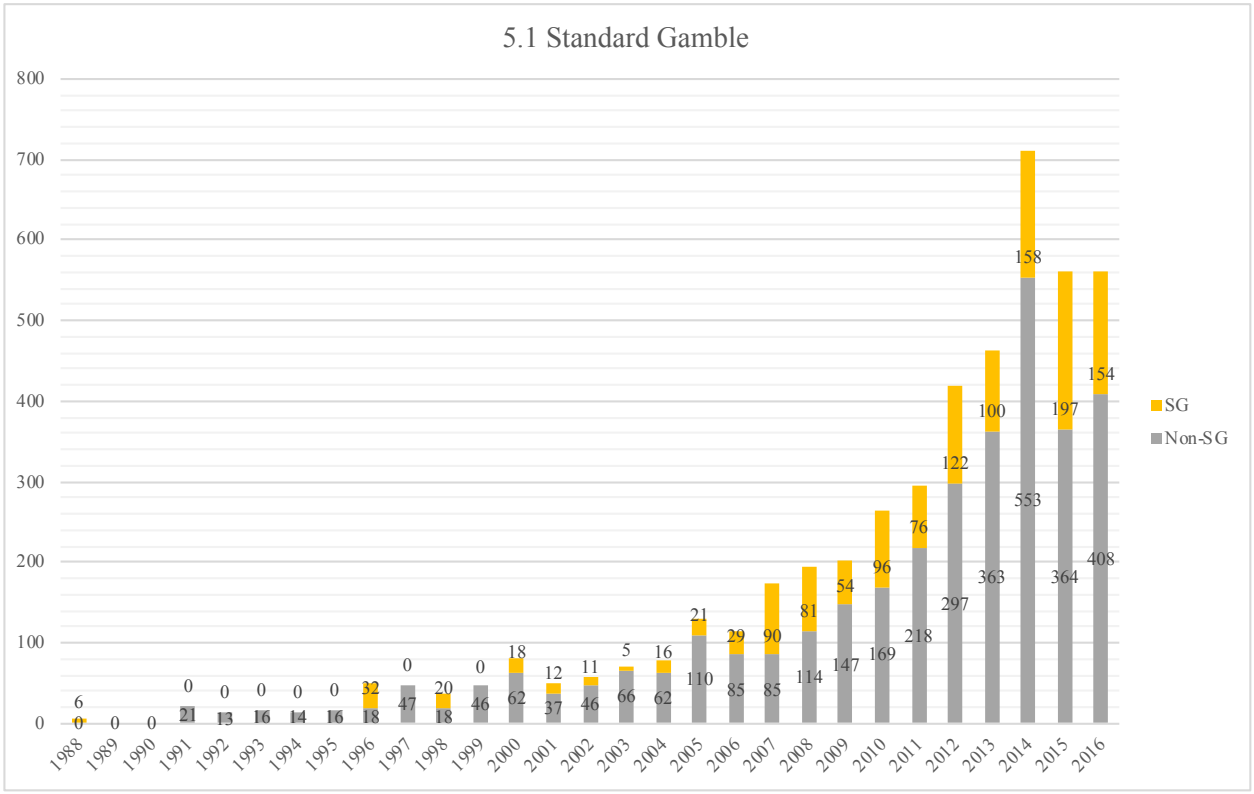
Figure 4. Measurement method classifications of the 4693 HUs used in oncology CUAs

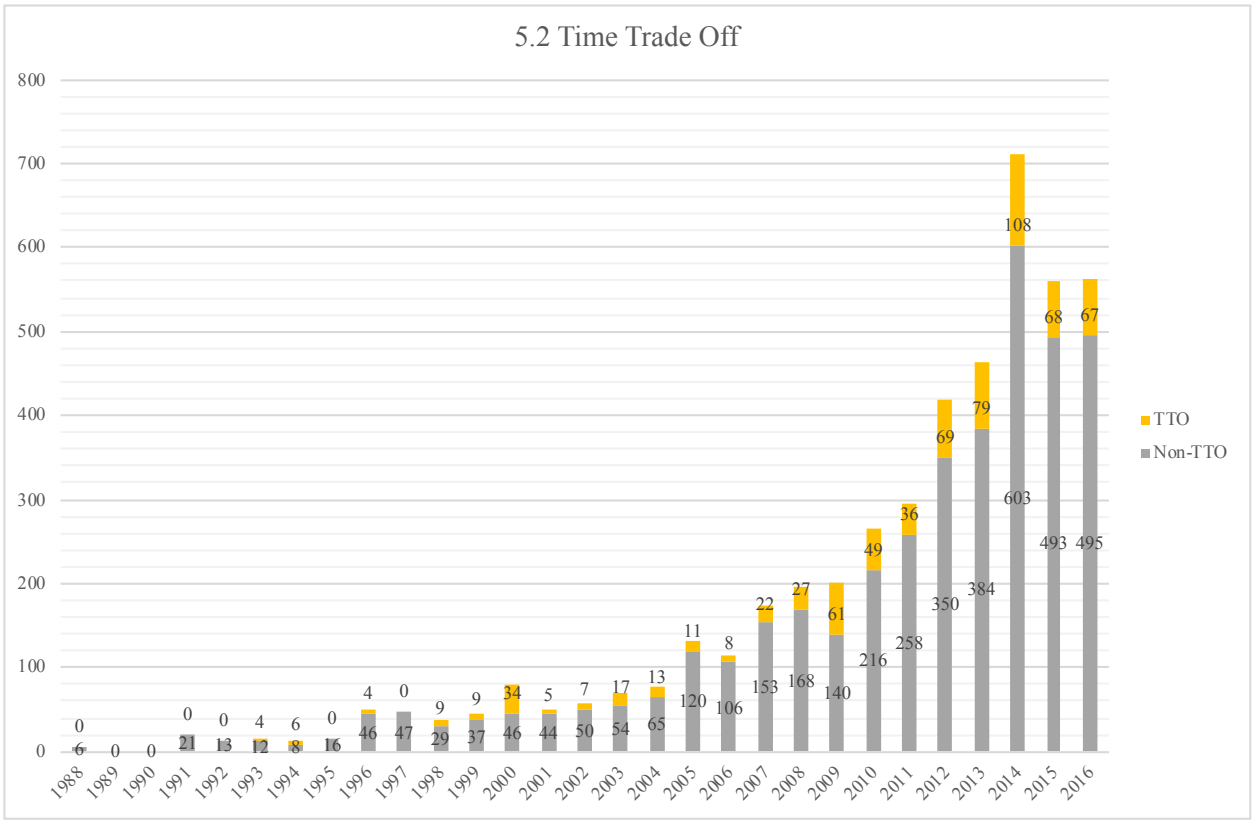
* The total percentage of all the methods is more than 1 because some HUs were measured by more than 1 method.

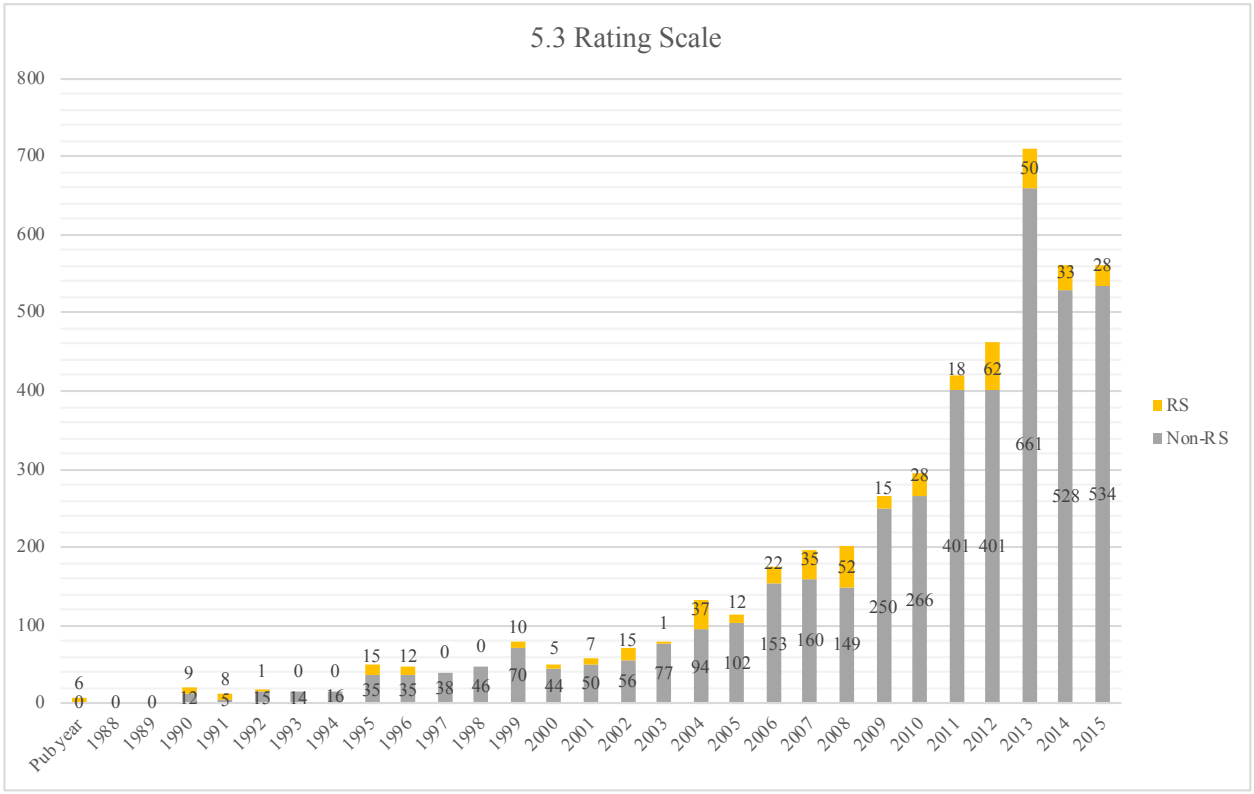
Besides, under indirect method classification, 327 (7.0%) HUs were first measured as HRQoL scores with instruments including Short-Form 12 (SF-12), Short-form 36 (SF-36), EORTC QLQ-C30, EORTC QLQ Head and Neck Module (EORTC QLQ-H&N35), and Functional Assessment of Cancer Therapy (FACT) questionnaires. These HRQoL data were multi-dimensional indexes that do not capture individual preference and could not be used in CUA directly. In this case, a mapping method, by transforming the HRQoL scores to a preference-based HU value by an established statistical algorithm, is required.[47~49] However, in fact, among those 327 HUs, only 204 of them (4.3% of 4693) were reported with a mapping method (i.e. mentioned in text and/or

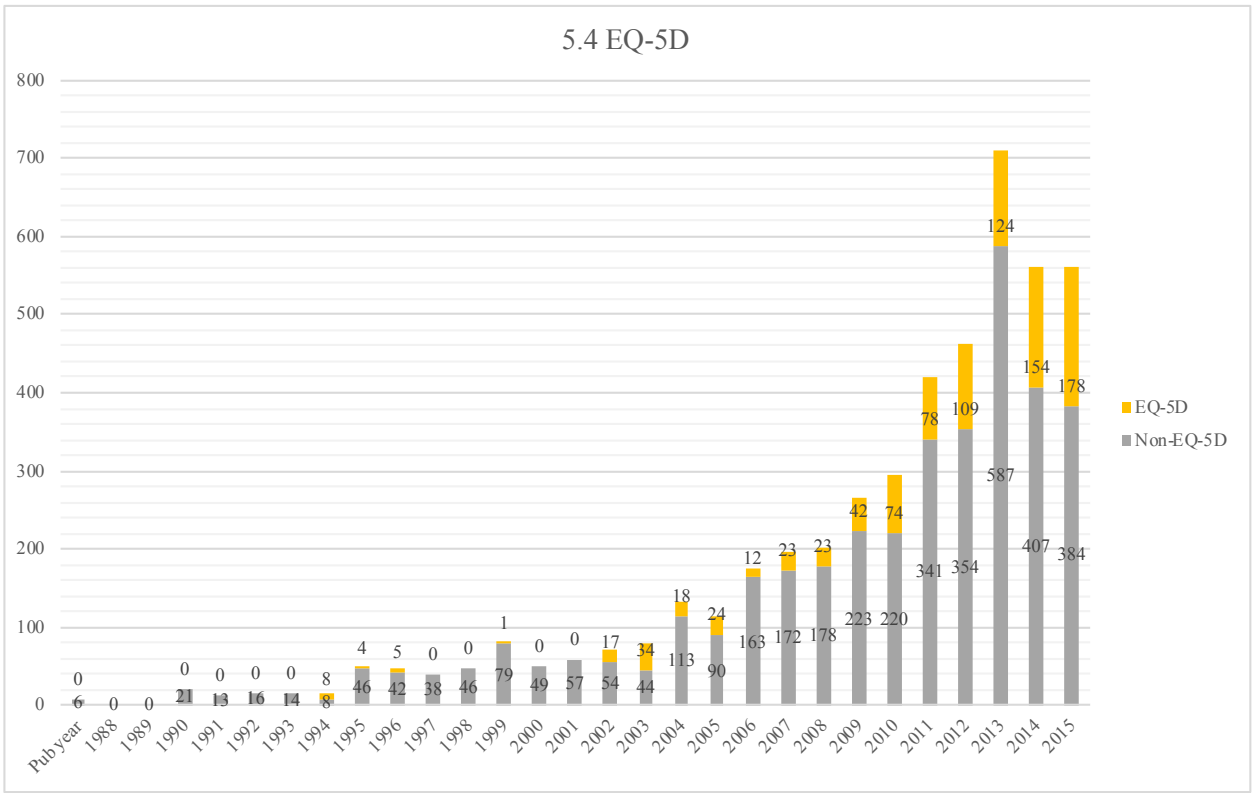
reference). At the same time, the other 123 (2.6%) HUs were used directly or transformed to HU indexes simply by dividing by the scale of QoL questionnaire (also called "linearly scale" in some CUAs). For example, a CUA of nivolumab for advanced renal cell carcinoma in 2018.[50]

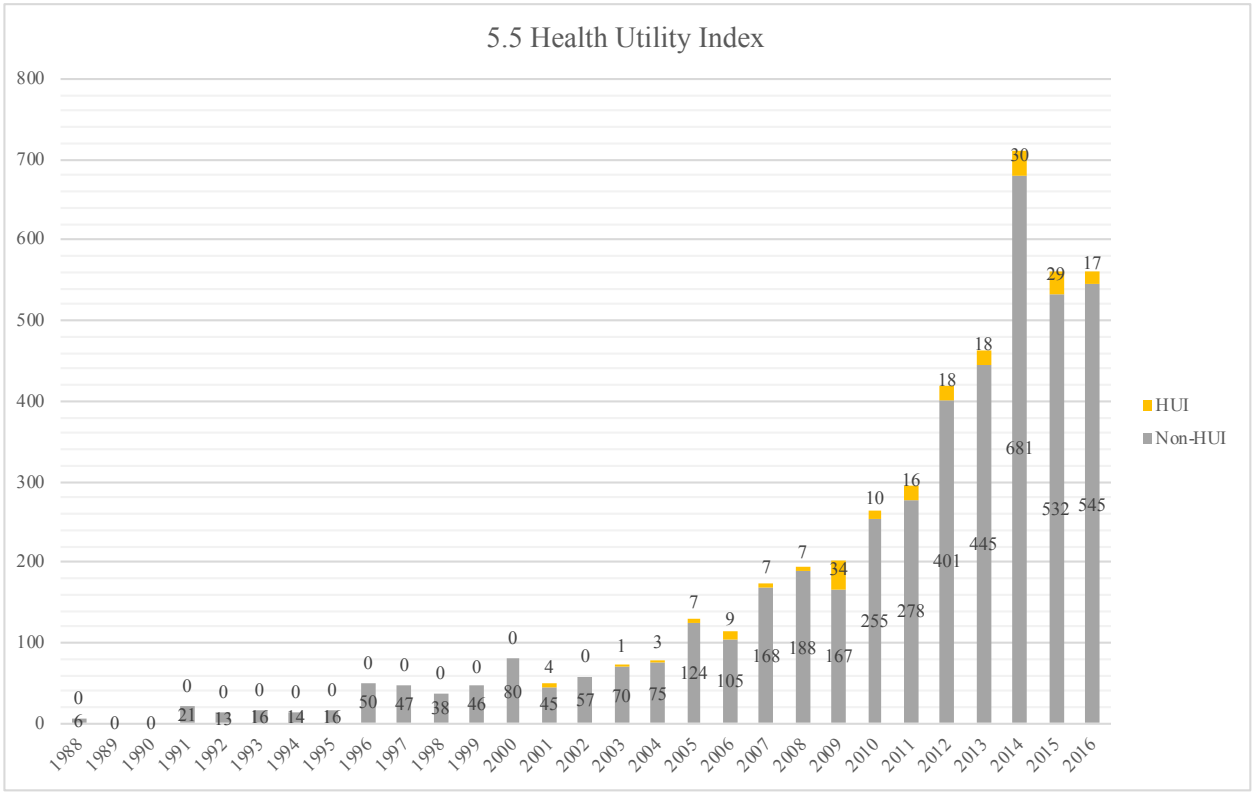
Furthermore, the change trends of HUs measured with 7 specified methods on a yearly basis are shown in Figures 5.1 to 5.7 (directly with TTO, SG, and RS techniques, indirectly with EQ-5D and HUI questionnaires, measured with HRQoL instruments first and those determined by expert opinions/author assumptions, respectively). The proportions of HUs measured with the SG and TTO methods tended to increase (Figures 5.1 and 5.2), while that measured by RS tended to decrease over time (Figure 5.3). In oncology CUAs published before 2000, very few HUs were measured by EQ-5D and none measured by HUI instruments. Afterwards, the proportions of HUs measured with EQ-5D and HUI instruments both tended to increase, especially for EQ-5D (Figures 5.4 and 5.5).

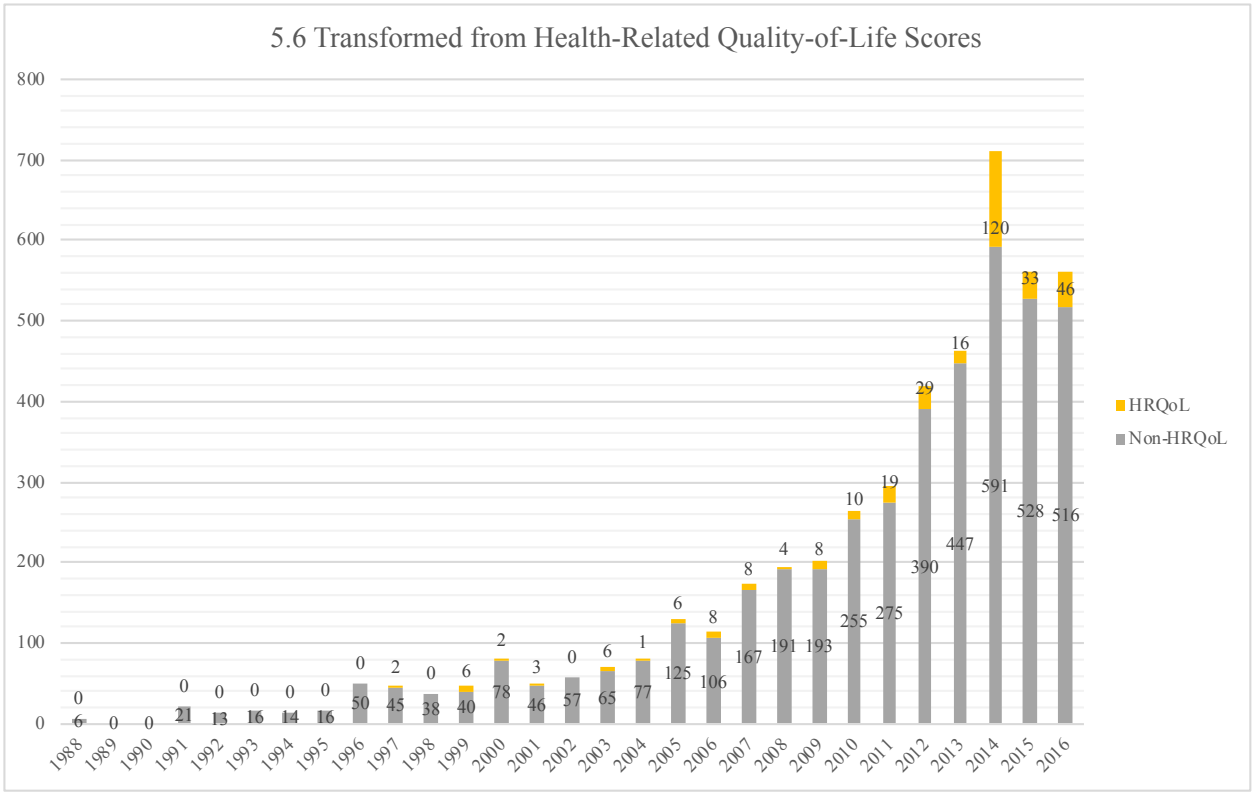

5.1 Standard Gamble
800
700
600
500
400
300
200
100
0
SG
Non-SG
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
0 0 0 0 0 0 0 0 32 0 20 0 18 12 11 5 16 21 29 90 81 54 96 76 122 100 158 197 154
6 0 0 21 13 16 14 16 18 47 18 46 62 37 46 66 62 110 85 85 114 147 169 218 297 363 553 364 408


5.2 Time Trade Off
800
700
600
500
400
300
200
100
0
TTO
Non-TTO
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
0 0 0 0 0 4 6 0 4 0 9 9 34 5 7 17 13 11 8 22 27 61 49 36 69 79 108 68 67
6 0 0 21 13 12 8 16 46 47 29 37 46 44 50 54 65 120 106 153 168 140 216 258 350 384 603 493 495


5.3 Rating Scale
800
700
600
500
400
300
200
100
0
Pub year
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
6 0 0 9 8 1 0 0 15 12 0 0 10 5 7 15 1 37 12 22 35 52 15 28 18 62 50 33 28
0 0 0 12 5 15 14 16 35 35 38 46 70 44 50 56 77 94 102 153 160 149 250 266 401 401 661 528 534
RS
Non-RS


5.4 EQ-5D
800
700
600
500
400
300
200
100
0
Pub year
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015
0 0 0 0 0 0 0 4 5 0 0 1 0 0 17 34 18 24 12 23 23 42 74 78 109 124 154 178
6 0 0 21 13 16 14 8 46 42 38 46 79 49 57 54 44 113 90 163 172 178 223 220 341 354 587 407 384
EQ-5D
Non-EQ-5D


5.5 Health Utility Index
800
700
600
500
400
300
200
100
0
HUI
Non-HUI
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
0 0 0 0 0 0 0 0 0 0 0 0 0 4 0 1 3 7 9 7 7 34 10 16 18 18 30 29 17
6 0 0 21 13 16 14 16 50 47 38 46 80 45 57 70 75 124 105 168 188 167 255 278 401 445 681 532 545


5.6 Transformed from Health-Related Quality-of-Life Scores
800
700
600
500
400
300
200
100
0
HRQoL
Non-HRQoL
1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
0 0 0 0 0 0 0 0 0 2 0 6 2 3 0 6 1 6 8 4 8 10 19 29 16 120 33 46
6 0 0 21 13 16 14 16 50 45 38 40 78 46 57 65 77 125 106 167 191 193 255 275 390 447 591 528 516

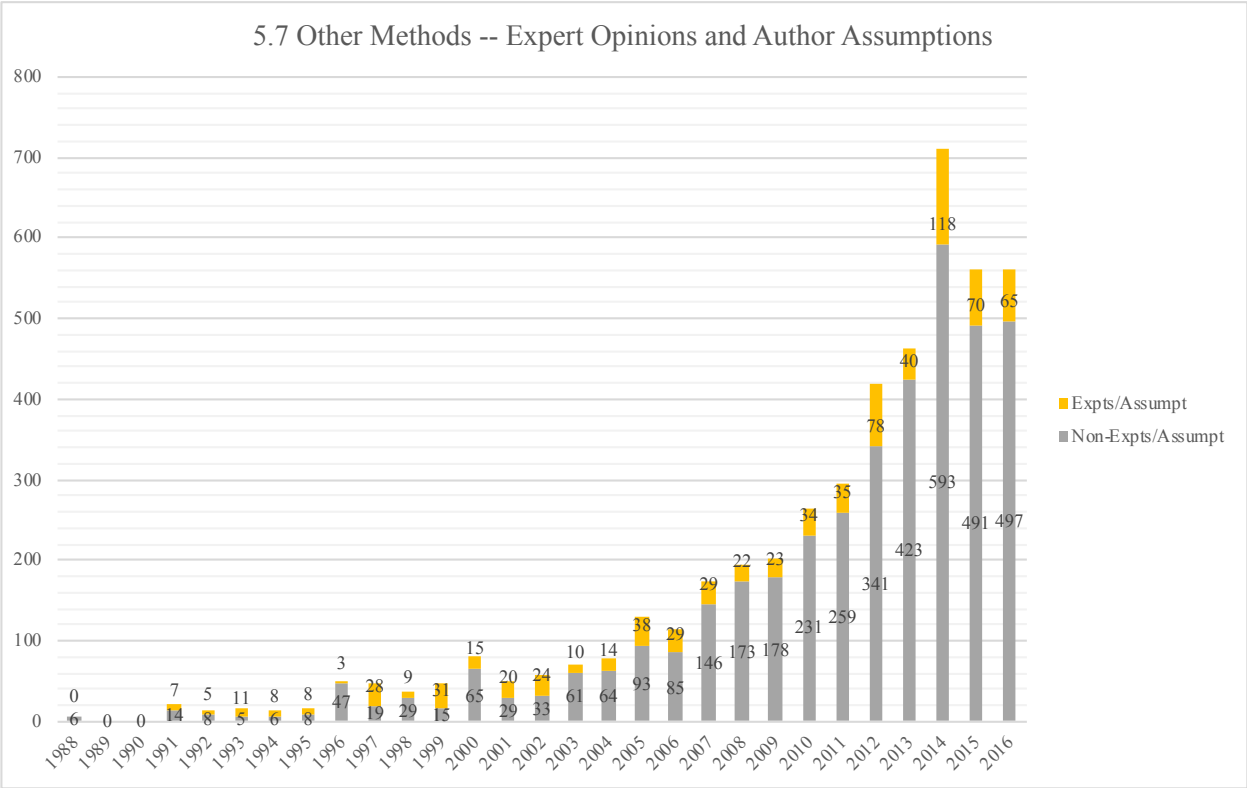


Figure 5. Change trends of 7 measurement methods of HUs used in oncology CUAs **Figure 5.1** SG **Figure 5.2** TTO **Figure 5.3** RS **Figure 5.4** EQ-5D **Figure 5.5** HUI **Figure 5.6** Transformed from QoL scores **Figure 5.7** Other methods (expert opinions and author assumptions)

* One HU could be measured by multiple methods.

Respondents of HUs When it comes to the subjects who rated HUs, patients (with the target cancer or related cancer disease) in oncology CUAs, accounted for 38.7% of the 4693 HUs (n=1816), making it the largest group among all the respondents. Samples from general population (also known as community-based samples) accounted for 22.5% (n=1055). Health professionals, including physicians, nurses, other medical staff and academic researchers, took up to 31.9% (n=1497). Among those health professionals who assigned HU values, about half of them (n=745, 15.9%) were measured with specified methods, while the other half (n=752, 16.0%) came from simply “expert opinions” or author assumptions. The 325 (6.9%) HUs left were judged being rated

by more than one kind of respondents, including 107 (2.3% of 4693) HUs derived from pool data in reviews of HUs (Table 6 and Figure 5).

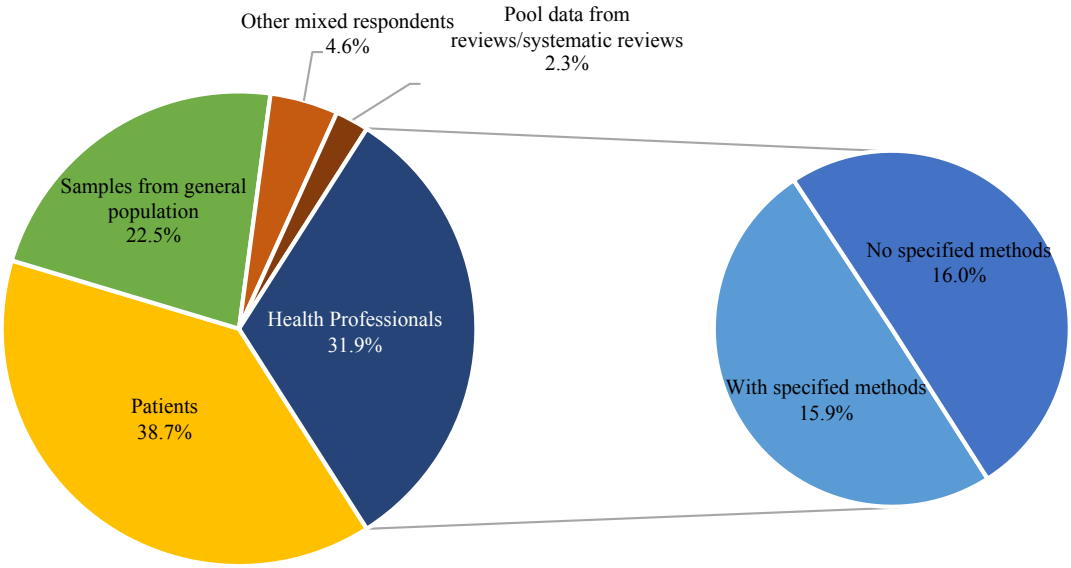


Figure 6. Respondents of the 4693 HUs used in oncology CUAs

Our outcomes also indicated that the proportion of HUs elicited from patients increased from 16.7% (1988-2000) to 37.2% (2001-2010) and to 41.9% (2011-2016), and the proportion of those from general population increased from 7.8% (1988-2000) to 21.3% (2001-2010) and to 24.7% (2011-2016). On the opposite, the proportion of HUs elicited from health professionals decreased from 71.8% (1988-2000) to 38.6% (2001-2010) and to 24.3% (2011-2016), making the remarkable reduction among all the subject types (Table 6).

Table 6. Characteristics of the 5583 HUs used in oncology CUAs

Characteristic classification	1988-2000 (n=355)	2001-2010 (n=1550)	2011-2016 (n=3678)	Total (n=5583)
Source of HUs [i]				
Primary HUs	244 (68.7%)	417 (26.9%)	692 (18.8%)	1353 (24.2%)
Secondary HUs from accessible original source (subtotal n=3360)				
Published review [ii]	2 (0.6%)	48 (3.1%)	176 (4.8%)	226 (4.0%)
Published individual study	101 (28.5%)	871 (56.2%)	2142 (58.2%)	3114 (55.8%)
Published study with non-related HUs[iii]	1 (0.3%)	2 (0.1%)	17 (0.5%)	20 (0.4%)
Secondary HUs from inaccessible original source (subtotal n=870)				
Unpublished source [iv]	0 (0)	48 (3.1%)	92 (2.5%)	140 (2.5%)
Wrong reference (study without HU)	4 (1.1%)	48 (3.1%)	144 (3.9%)	196 (3.5%)
No/incomplete reference [v]	0 (0)	63 (4.1%)	149 (4.1%)	212 (3.8%)
Other inaccessible sources [vi]	3 (0.8%)	53 (3.4%)	266 (7.2%)	322 (5.8%)
Reference level of secondary HUs (for published reviews and individual studies, subtotal n=3340)				
1st level	98 (95.1%)	729 (79.3%)	1779 (76.7%)	2606 (78.0%)
2nd level	4 (3.9%)	182 (19.8%)	500 (21.6%)	686 (20.5%)
3rd or 4th level	1 (1.0%)	8 (0.9%)	39 (1.7%)	48 (1.4%)
Measurement method (for primary HUs and secondary ones from accessible reviews/individual studies, subtotal n=4693)				
Direct measurement	195 (56.2%)	732 (54.8%)	1376 (45.7%)	2303 (49.1%)
Indirect measurement	27 (7.8%)	339 (25.4%)	1170 (38.9%)	1536 (32.8%)
Other methods (expert opinions and author assumptions)	125 (36.0%)	242 (18.1%)	405 (13.5%)	772 (16.5%)
Multiple methods [vii]	0 (0)	14 (1.0%)	35 (1.2%)	49 (1.0%)
Unclear	0 (0)	9 (0.7%)	24 (0.8%)	33 (0.7%)
HU respondent (for primary HUs and secondary ones from accessible reviews/individual studies, subtotal n=4693)				
Cancer patients	58 (16.7%)	497 (37.2%)	1261 (41.9%)	1816 (38.7%)
Samples from general population	27 (7.8%)	284 (21.3%)	744 (24.7%)	1055 (22.5%)
Health professionals [viii]	249 (71.8%)	517 (38.6%)	732 (24.3%%)	1497 (31.9%)
Mixed respondents [ix]	13 (3.7%)	39 (2.9%)	273 (9.1%)	325 (6.9%)

CUA cost-utility analysis; HU health utility

[i] The publication years of the extracted HUs were recorded according to the oncology CUAs in which they were used.

[ii] "Published review" refers to systematic reviews and reviews of HUs, with reference level equal to 2

[iii] "Non-related HUs" refer to HUs in referenced studies that could not be matched to the extracted ones, for either health states or utility values, and disabled reviewers to locate potential original HUs.

[vi] "Unpublished source" refers to the CUA that provided the HU for the first time, while saying that it was measured in a separate study (e.g. a phase III trial). In this case, the CUA was not a primary study.

[v] Incomplete reference included a last name and a year only, or the reference of related database only. In this case, the original HU studies were inaccessible due to lack of information. The referenced databases consisted of the CEVR CEA Registry[33] and the CEA Registry by Harvard School of Public Health [51].

[vi] "Other inaccessible sources" refer to conference proceedings, books, drug manufacturer-provided data, papers not in English and so on.

[vii] "Multiple methods" refer to 2 or more measurement method types (among direct, indirect or other methods).

[viii] "Health professionals" includes physicians, nurses, other medical staffs, researchers in health science and so on.

[ix] "Mixed respondents" refer to pool data in (systematic) reviews and samples from more than one type of respondents above.

Consistency of using HUs

Secondary HUs tracked to accessible studies, including those without health-utility-related data, were included for the consistency assessment (subtotal n=3360). In general, the consistency of HUs between the oncology CUAs and original HU studies were classified and assessed based on both the two components of a HU parameter— the health state description and utility value.

According to our findings (Table 7 and Figure 7), 1348 of the 3360 secondary HUs (40.1%) derived from accessible sources in oncology CUAs were the same in both description and value as those in original HU studies. This number decreased from 68.3% (1988-2000) to 33.6% (2001-2010) and then increased slightly to 41.5% (2011-2016). 633 (18.9%) and 390 (11.6%) differed only in value and description, respectively, compared with the original studies. 989 (29.4%) had both the descriptions and values different from those in original studies. This number increased from 8.7% (1988-2000) to 34.0% (2001-2010) and then slightly decreased to 28.6% (2011-2010).

Among the 633 HUs with different value only, 132 (20.9%) were derived from the utility decrements (i.e. disutilities) or increments from original HU studies, or vise versa.

Out of those HUs with different description only (n=390), 247 (63.3%) were derived from same disease, but different severity and/or history (n=100, 25.6%), different interventions, adverse events or comorbidities (n=126, 32.3%), different respondent demographics (n=17, 4.4%), or no explanation (n=4, 1.0%). 143 (36.7%) were derived from different diseases, due to adverse events, comorbidities or other related health states that could happen in target patients (n=87, 22.3%), or with no explanation (n=56, 14.4%) (Table 7).

Among all the HUs with either or both description and value different from those in original studies, about 377 out of 2012 (18.7%) of them were different, but without explanation (Table 7).

Table 7 Consistency of HUs between oncology CUAs and original HU studies

Component	1988-2000	2001-2010	2011-2016	Total
All HUs	**n=104**	**n=921**	**n=2335**	**n=3360**
Same description and value	71 (68.3%)	309 (33.6%)	968 (41.5%)	1348 (40.1%)
Different value only	10 (9.6%)	158 (17.2%)	465 (19.9%)	633 (18.9%)
Different description only	14 (13.5%)	141 (15.3%)	235 (10.1%)	390 (11.6%)
Different description and value	9 (8.7%)	313 (34.0%)	667 (28.6%)	989 (29.4%)
Different value only	**n=10**	**n=158**	**n=465**	**n=633**
For utility increment/disutility	0 (0)	23 (14.6%)	109 (23.4%)	132 (20.9%)
With other explanation	4 (40.0%)	114 (72.2%)	292 (62.8%)	410 (64.8%)
Without explanation	6 (60.0%)	21 (13.3%)	64 (13.8%)	91 (14.4%)
Different description only	**n=14**	**n=141**	**n=235**	**n=390**
Same disease, with justification for				
Different severity/history	4 (28.6%)	40 (28.4%)	56 (23.8%)	100 (25.6%)
Different interventions/AEs/ comorbidities	6 (42.9%)	53 (37.6%)	67 (28.5%)	126 (32.3%)
Different patient demographics	0 (0)	11 (7.8%)	6 (2.6%)	17 (4.4%)
No explanation	0 (0)	4 (2.8%)	0 (0)	4 (1.0%)
Different disease, with justification for				
AE/comorbidity/other related HS	2 (14.3%)	13 (9.2%)	72 (30.6%)	87 (22.3%)
No explanation	2 (14.3%)	20 (14.2%)	34 (14.5%)	56 (14.4%)
Different description and value	**n=9**	**n=313**	**n=667**	**n=989**
With explanation	5 (55.6%)	208 (66.5%)	550 (82.5%)	763 (77.1%)
Without explanation	4 (44.4%)	105 (33.5%)	117 (17.5%)	226 (22.9%)

AE adverse event, HS health state

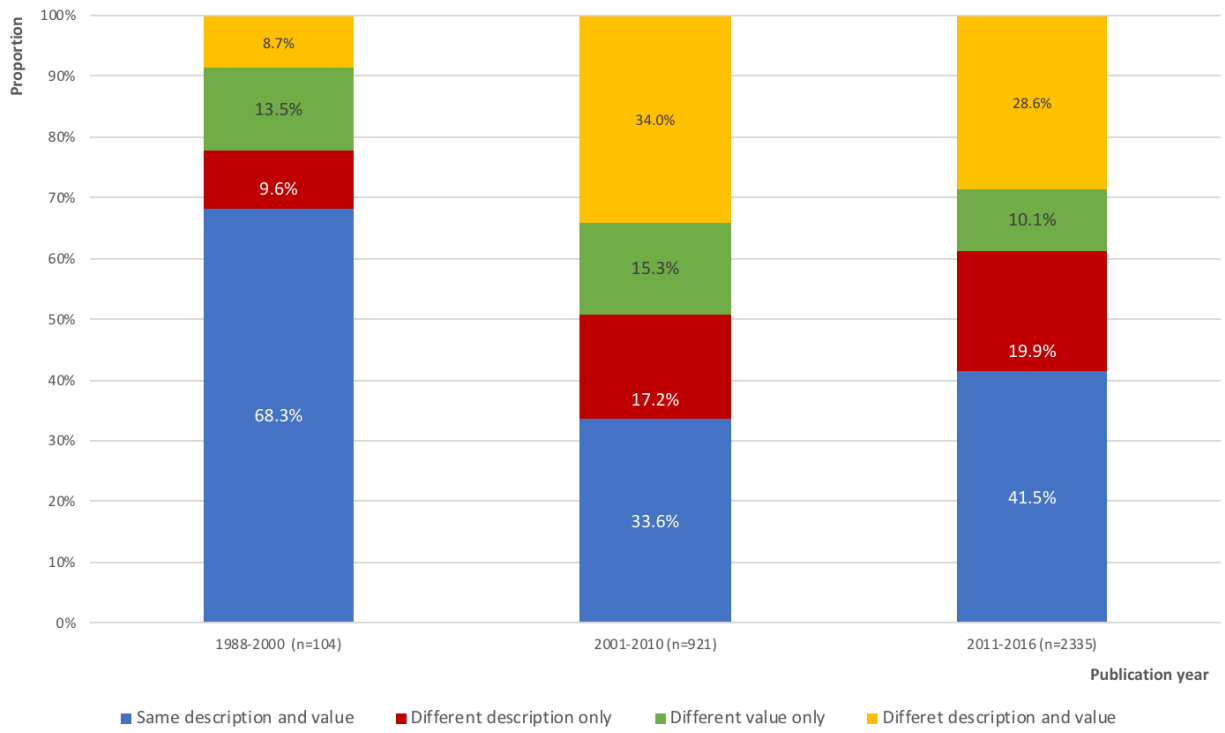


Figure 7. Change of consistency in the 3360 HUs overtime (1988-2016)

Discussion and recommendations

This chapter aimed to review the basic characteristics and assess the consistency of the HUs applied in oncology CUAs published between 1976 and 2016, based on the CEVR CEA Registry. Overall, the number of the oncology CUAs between 2011 and 2016 was about 10 times of that between 1988 and 2000, the same pattern existed for the number of extracted HUs at the same time. As time passed by, keep using HU data from old original HU sources published decades ago (Figure 3) seems to be questionable.

Our findings indicated that most of the secondary HUs used in oncology CUAs (2606 out of 3340, 78.0%) were cited directly from original sources, whenever accessible. Overall, about 50% of the HUs used were measured directly, while the other half measured by indirect (34.2%) or other methods (16.5%). Cancer patient is the largest kind of respondents among all the 4693 accessible

HUs, accounting for about 38.7%, followed by health professionals (31.9%) and samples from general population (22.5%).

The application of indirect methods (e.g. EQ-5D and HUI) were barely used before 2000, probably because the relatively late development of these preference-weighted HRQoL questionnairs compared with direct methods – the SG and TTO were first developed in the 1950s and the 1970s,[12, 15] while the EQ-5D and HUI were first developed in the 1990s and 1980s,[17, 18] respectively.

The proportion of HUs measured by SG, TTO technique and EQ-5D questionnaire tended to increase gradually in oncology CUAs from 1988 to 2016. In the same period, the proportion of HUs came from health professionals without any specified measurement method in oncology CUAs tended to decrease steadily. Those came from patients and community-based samples tended to increase fast, on the contrary. These practice trends are similar to those from the previous review of published CUAs in 2006 [39]. All these findings indicate the remarkable development of HU principles, sources and applications in the cancer-related cost-effectiveness field.

Overall consistency results

For the consistency of using HUs between oncology CUAs and original sources, first, the proportion of the same HUs from original HU studies (40%) might be overestimated. Because some difference was found being made without reasonable explanations during reviewing process, and the rightness of all those adjustments made was relatively subjective and beyond the scope of our study.

18.7% (377 out of 2012) accessible secondary parameters different from original studies were used without any explanation. Furthermore, the estimation of some different HUs with explanation were

questionable. For example, 36.7% (n=143) of the 390 secondary health states with different values only were derived from different disease (e.g. heart disease or hip fracture) due to common side effects, comorbidities, other related health states in target cancer patients or no explained reason. It is a questionable change of applying a HU weight from other disease on certain health condition in cancer patients by simply changing the health state description. Such result is probably due to lack of ideal cancer-related HUs. This finding reflects the needs of future studies measuring a series of cancer-related HUs or other HRQoL scores in the same context, especially HUs for common side effects and comorbidities in cancer patients. Besides, more studies and guidance of a reliable HU adjustment process are welcome as well.

Additional to the inconsistency and questionable adjustments in some HU used in oncology CUAs, several other issues exist between the ideal methodology and real-world practice.

Reporting issue: lack of transparency and accuracy

The accuracy and transparency of reporting in published oncology CUAs needs to be further improved. First, for referencing, 3.5% of all the 5583 HUs (n=196) were tracked to studies with no HU data, and 3.8% of them (n=212) had no reference or incomplete references that made readers unable to find the source. Among the HUs that were tracked to accessible HU studies, 22.0% (734 out of 3360) cited the original studies indirectly. Instead of increasing over time as we expected, the proportion of direct references (1st level reference in our findings) actually decreased from 95.1% (1988-2000) to 79.3% (2001-2010) and to 76.7% (2010-2016).

Furthermore, when the oncology CUAs referenced HUs indirectly from non-original studies (e.g. previous CUAs using secondary HUs), the health state descriptions and/or utility values were usually adjusted in the referenced studies to fit their contexts. Thus, it is highly possible that the indirect referencing of HUs was inaccurate and inappropriate.

For example, the CUA by Janice et al.[52] investigated treatments for stage I and II endometrial cancer. Corresponding HU parameters were cited from a published CUA in 2002.[53] However, the cited HUs were actually originally measured in another study [54] (the original HU study 1988) and were already adjusted in the referenced CUA (Table 7). Beside the HUs, information about the population was different as well. Given the difference between the demographics of the respondents in the original study and the target cohort in the example CUA, the HU parameter "stage II endometrial cancer" (with a utility value equal to 0.56) and its variance lack validity and explanation, even if it was tested widely in SAs. The health state "stage I endometrial cancer" was different from that in the original study too. The cost-effectiveness outcome based on this input parameters lacked robustness as a result.

Table 8. Example of the inaccuracy of HUs results from indirect reference in an oncology CUA

Study and Pub year	Example CUA 2007 [52]	Reference CUA 2002 [53]	Original HU study 1998 [54]
Target Disease	Endometrial cancer (stage I and II)	Atypical squamous cells of undetermined significance	Chronic conditions, including cancers [i]
Population	Hypothetical women diagnosed at the age of 63 in Ontario CA.	Hypothetical adolescent girls entered the model at age 13 years in US.	The US general females with genital cancer, n=27, mean age =39.
Health state description	i) Stage I endometrial cancer ii) Stage II endometrial cancer	i) Base case: local cancer ii) Regional cancer	Female genital cancer
Utility value (plausible range)[ii]	i) 0.68 (0.50-0.80) ii) 0.56 (0.40-0.70)	i) 0.68 (0.60-1.00) ii) 0.56 (0.40-1.00)	0.68 (interquartile 0.48-0.84)

HS Health State HU health utility
[i] Based on NHIS (the National Health Interview Survey) data.
[ii] The plausible ranges in 2 CUAs refer to the range tested in SAs.

Additionally, cases of inaccurate and/or non-transparent reporting were also found in oncology CUAs which cited 1st-level HU references. For example, in a CUA of detection technology for oesophageal or gastric cardia cancer [55], the authors built a decision analytic model with 4 health states and stated that "life expectancy and QALYs are taken from a previous study"[56]. The estimated health outcomes (LYs and QALYs) associated with each health state were reported right afterwards without their corresponding utility values. The process of this economic analysis was highly non-transparent.

Take the study by S. Abellea et al.[57] as another example. Beside the 2 HUs derived from literature, the authors in this study mentioned the HU of "a disease-free survivor" and utility decrements due to chemotherapy-induced toxicities from other published studies. However, neither the health

states nor corresponding utilities were stated clearly. As a result, the application of HU parameters was obscure and only part of the HU input was extracted from the study.

Second, for the reporting of measurement methods of HUs in oncology CUAs, chances were that the measurement method(s) of the secondary HU was not specified, while there were multiple HU weights from more than one measurement methods in the original study(ies). At the same time, the utility values in the oncology CUA was different from those in the original source(s). In this case, the information in the oncology CUA was insufficient for readers to judge how the parameter was originally measured and adjusted, if needed. As a result, the reporting in such oncology CUAs was inaccurate and lacked basic information.

Simply "borrowing" a HU from a non-primary study, with or without further adjustment, is not a reliable way to perform a CUA. For the purpose of transparency and accuracy in reporting, CUA authors should at least briefly introduce the population and measurement method of the HUs used based on the original HU study(ies). The reference(s) should be recorded directly and accurately, accompanied with clear explanations if needed. A lack of either could lead to information loss and/or confusion and prevent cost-effective results from being used.

Misunderstanding the concept of HU

HRQoL and HU have different definitions. HRQoL is a multi-dimensional concept incorporating related factors, including physical and emotional factors, while the HU score is a cardinal index for measuring HRQoL under uncertainty, on the 0-1 scale.[58,59] HU is one of many approaches that measures HRQoL numerically. An HU value, represents the strength of the individuals' preferences for alternative health outcomes.[58~60] HUs can be measured, directly or indirectly, by a

number of techniques and instruments. HEE researchers can also convert them from HRQoL data using mapping algorithms when there is no ideal HU data available. The application of HRQoL-score-derived HUs in published CUAs in oncology, however, was problematic.

Our findings (Figure 3 and text) show that 37.6% of the HRQoL-derived HUs (123 out of 327) were transformed improperly to HU weights (by "linearly scale"), or were used directly with no mapping process at all.

Take a CUA of stage III ovarian cancer treatments as an example. The four HU parameters used in this study were first measured with the FACT-Ovarian (FACT-O) questionnaire in a previous study, then transformed to "quality of well-being index values" by directly dividing by a maximum FACT-O score of 156 in the CUA. [61, 62] Two of the HU parameters were further adjusted based on QoL increasement, measured by the EORTC QLQ-C30 questionnaire from another study.[63] This whole estimation process reflected a confusion between HRQoL scores and HUs and partial understanding of QALY. Other cases of misusing HUs also include generating HU weights from event rates, or even usefulness of certain health service (e.g. patients' preference of hospice services[64, 65]). None of the numbers described above was a proper HU application.

Given the conditions described above, some of which still could be found in the last decade, the understanding and applying of HU remained to be problematic. We suggest that special attention should be paid when researchers try to transform HRQoL data to obtain HU scores or cite HUs that were converted from HRQoL scores in other studies. An established mapping function and proper referencing are needed. Furthermore, the mapping method should be validated with out-of-sample data in order to assure the validity and reliability of the method [48]. Guidance about selecting mapping methods to derive HUs from HRQoL data is expected in the future.

Difference between HU values measured by different methods

First, although the CADTH guideline in 2009 recommends using SG or TTO techniques when HUs are elicited directly,[27] about 10% of the HUs used in oncology CUAs were elicited by the RS. An RS is generally easy for non-professional subjects to understand; however, it measures HUs without uncertainty. Thus, the RS produces non-utility values and is not an ideal rating approach, if assessed strictly.

Second, on average, it was estimated that 36.5% (333 out of 912) of oncology CUAs used HUs elicited by more than 1 specified method. The number of this kind of studies grew gradually as time passed by (Figure 9). This is probably resulted from the availability of more published cancer-related HU studies, as well as the unclear reporting of the measurement methods of HUs used in some oncology CUAs. When there was more than one available measurement method used in the same health state, CUA authors usually simply chose the mean or median estimates from multiple available utility values as their HU input. However, the difference between utility values for the same health state resulting from different measurement methods does exist and could lead to bias if not handled properly.[59, 66]

Of note, for the same health state, the SG score is usually higher than the TTO score.[66] Additionally, the utility values measured by direct methods tend to be larger than those measured by indirect methods.[59] Several previous studies also advised that the uncertainty of a HRQoL-derived HU could be underestimated (i.e. a narrower CI).[59, 67,68] As a result, we suggest that researchers try to use all HUs measured by the same method in one CUA,[59] or at least use all the HU inputs measured either directly or indirectly. Otherwise, a reasonable adjustment involving all the eligible HU values is needed.

Means versus Medians

Among HU studies, some provide median estimates, some provide means, and some report both. Some authors stated that the medians are less sensitive to extreme scores, so that they report the median utility values as the primary outcome,[69] while the mean estimates of HUs were found to be the most commonly recommended and used input parameters for CUAs. Furthermore, chances were that some CUA authors cited both median and mean HU values from literature in one study. Guidance and consensus about the choice between mean and median HU estimates for CUA are needed.

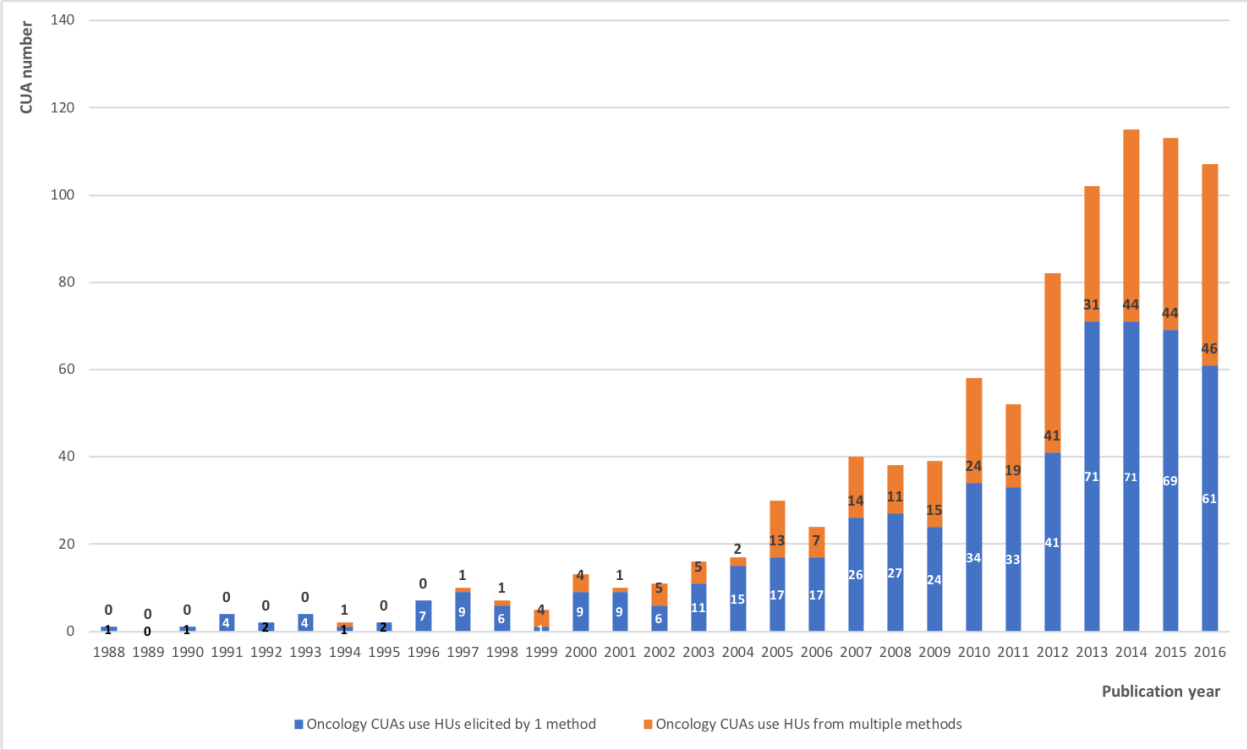


Figure 8. Oncology CUAs that used HUs from multiple measurement methods (study n=912)

Conclusion

The increased number of HUs and corresponding references per study indicated that the application of HUs in published oncology CUAs gradually became more complex over time. The measurement methods and respondent sampling of HUs used in oncology CUAs tended to be relatively less arbitrary, given the increased use of HUs measured by SG, TTO or EQ-5D, and the deceased use of HUs derived from simply "expert opinions" or author assumptions.

The consistency assessment showed that about 19% of HUs with either or both descriptions and values different from those in original studies without any explanation. Approximately 7% with different descriptions only were derived from different diseases, the estimation of which may be questionable.

Some other methodology issues were also found during the review, including the inaccurate and non-transparent reporting methods, misunderstanding of the concept of HU and the mixed use of HUs measured with different methods without proper adjustment.

CHAPTER IV. CONCLUSION

Limitations

The study has some limitations. The collection of cancer-related HUs was not comprehensive. This study only identified eligible oncology CUAs published in English and were indexed in MEDLINE (according to the CEVR CEA Registry). It is possible that some eligible CUAs included in other databases were missing. CUAs published after 2016 were not included due to time limit. Besides, the data extraction processes were not performed in duplicate. In order to control the internal data validity, the corresponding HU information from the registry database was used as references. Although the data extraction bias may not be avoided completely, considering the large volume of studies included, we suppose that the features and issues found are representative in published oncology CUAs and HUs used. Third, we did not assess the quality of original HU studies. It was also beyond the scope of our study to judge if the changes and explanations that CUA authors made for the health utilities were reasonable or not.

Conclusions

In total, 912 oncology CUAs published between 1988 and 2016 and 5583 HUs used in them were identified and reviewed in this registry-based study.

Overall, the increasing interest and development of conducting oncology CUAs and application of related HUs were obvious, especially after 2000. At study level, the majority of oncology CUAs used the model-based approach, adopted the health care payer's perspective, and reported ICURs while many lacked the use of life-time horizon and sufficient SA related to HUs.

At individual HU level, the consistency of the secondary HUs with accessible original studies remained problematic. About 19% of the secondary HUs with different description and/or values were used in CUAs without explanation, and 7% of those secondary HUs different from original ones were derived from different diseases. The majority of cancer-related HUs were derived from literature with original sources referenced in CUAs of cancer disease. The most commonly used direct and indirect measurement methods were SG and EQ-5D, respectively. Patients were the most commonly involved respondents. Beside the consistency problem, the other issues found included non-transparent reporting, inaccurate references, misunderstanding of the concept of HU, and lack of reasonable adjustment for the weighted utility values derived from multiple measurement methods. We also highlighted the continued need for studies investigating adjustment methods for those HUs of cancer-related side effects and/or comorbidities derived from original studies of different diseases.

Some method issues like non-transparent and/or inaccurate reporting and misunderstanding of HUs could be avoided completely if researchers pay enough attention, while the others need further investigation and explicit guidance.

References

1 National Cancer Institute. What is Cancer. Feb 2015. Available from URL: https://www.cancer.gov/about-cancer/understanding/what-is-cancer. [accessed on Nov 2018].

2 National Cancer Institute. Cancer Statistics. Apr 2018. Available from URL: https://www.cancer.gov/about-cancer/understanding/statistics. [accessed on Nov 2018].

3 World Health Organization. Global Health Observatory. Geneva: World Health Organization; 2018. Available from URL: http://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf. [accessed on Sept 2018].

4 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018 Sep 12.

5 Canadian Cancer Society. Media backgrounder: Canadian Cancer Statistics: A 2018 special report. June 2018. Avaiable from URL: http://www.cancer.ca/en/about-us/for-media/media-releases/national/2018/media-backgrounder-canadian-cancer-statistics-a-2018-special-report/?region=on [accessed on Aug 2019].

6 The official website of the Government of Canada. Publications-Diseases and conditions. Available from URL: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/fact-sheet-cancer-canada.html [accessed on Aug 2019].

7 de Oliveira C, Weir S, Rangrej J, Krahn MD, Mittmann N, Hoch JS, Chan KK, Peacock S. The economic burden of cancer care in Canada: a population-based cost study. CMAJ Open. 2018 Jan 4;6(1):E1-0.

8 Hopkins RB, Goeree R, Longo CJ. Estimating the national wage loss from cancer in Canada. Current Oncology. 2010 Apr;17(2):40.

9 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015 Sep 24.

10 Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. Journal of health economics. 1996 Apr 1;15(2):209-31.

11 Torrance GW, Feeny D. Utilities and quality-adjusted life years. International journal of technology assessment in health care. 1989 Oct;5(4):559-75.

12 von Neumann J, Morgenstern O. Theory of games and economic behavior. New York, Wiley, 1953.

13 Torrance GW. Measurement of health state utilities for economic appraisal: a review. Journal of health economics. 1986 Mar 1;5(1):1-30.

14 Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. Journal of clinical epidemiology. 2003 Apr 1;56(4):317-25.

15 Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health services research. 1972;7(2):118.

16 Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health services research. 1976;11(4):478.

17 Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. Operations research. 1982 Dec;30(6):1043-69.

18 Group TE. EuroQol-a new facility for the measurement of health-related quality of life. Health policy. 1990 Dec 1;16(3):199-208.

19 Schackman BR, Gold HT, Stone PW, Neumann PJ. How often do sensitivity analyses for economic parameters change cost-utility analysis conclusions?. Pharmacoeconomics. 2004 Apr 1;22(5):293-300.

20 McDonough CM, Tosteson AN. Measuring preferences for cost-utility analysis. Pharmacoeconomics. 2007 Feb 1;25(2):93-106.

21 Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, ISPOR Health Economic Evaluation Publication Guidelines-CHEERS Good Reporting Practices Task Force. 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. 2013 Mar 1;16(2):231-50.

22 Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment. Pharmacoeconomics. 2006 Apr 1;24(4):355-71.

23 Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. Bmj. 2011 Apr 11;342:d1766.

24 Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan T. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health economics. 2005 Apr;14(4):339-47.

25 Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Medical Decision Making. 2007 Jul;27(4):491-9.

26 Edwards RT, Charles JM, Lloyd-Williams H. Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. BMC public health. 2013 Dec;13(1):1001.

27 Mittmann N., Evans W.K., Rocchi A., Longo C. J., Au H.-J., Husereau D., Leighl N., Isogai P., Krahn M., Peacock S., Marshall D., Coyle D., Malfair Taylor S.C., Jacobs P., Oh P.I. Addendum to CADTH's Guidelines for the Economic Evaluation of Health Technologies: Specific Guidance for Oncology Products. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

28 PausJenssen AM, Singer PA, Detsky AS. Ontario's Formulary Committee. Pharmacoeconomics. 2003 Mar 1;21(4):285-94.

29 Buxton MJ. Economic evaluation and decision making in the UK. Pharmacoeconomics. 2006 Nov 1;24(11):1133-42.

30 Yong JH, Beca J, Hoch JS. The evaluation and use of economic evidence to inform cancer drug reimbursement decisions in Canada. Pharmacoeconomics. 2013 Mar 1;31(3):229-36.

31 Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Medical Decision Making. 2007 Jul;27(4):491-9.

32 Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: a progress report. Annals of Internal Medicine. 1987 Feb 1;106(2):275-91.

33 Center for the Evaluation of Value and Risk in Health. The Cost-Effectiveness Analysis Registry [Internet]. (Boston), Institute for Clinical Research and Health Policy Studies, Tufts Medical Center. Available from URL: www.cearegistry.org [accessed on Apr 2018].

34 StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP. Available from www.stata.com.

35 Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. Value in Health. 2006 Jul;9(4):213-8.

36 Noone AM, Howlader N, Krapcho M, et al. (eds). SEER Cancer Statistics Review, 1975-2015, National Cancer Institute, Bethesda, MD, http://seer.cancer.gov/ csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER website April 2018.

37 American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society; 2019.

38 de Lima Lopes Jr G, Segel JE, Tan DS, Do YK, Mok T, Finkelstein EA. Cost-effectiveness of epidermal growth factor receptor mutation testing and first-line treatment with gefitinib for patients with advanced adenocarcinoma of the lung. Cancer. 2012 Feb 15;118(4):1032-9.

39 Pandharipande PV, Harisinghani MG, Ozanne EM, Specht MC, Hur C, Lee JM, Gazelle GS. Staging MR lymphangiography of the axilla for early breast cancer: cost-effectiveness analysis. American Journal of Roentgenology. 2008 Nov;191(5):1308-19.

40 National Institute for Health and Care Excellence (NICE): Guide to the methods of technology appraisal. Available from URL: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 [accessed on May 2019].

41 Ara R, Brazier J, Zouraq IA. The use of health state utility values in decision models. Pharmacoeconomics. 2017 Dec 1;35(1):77-88.

42 Arnold D, Girling A, Stevens A, Lilford R. Comparison of direct and indirect methods of estimating health state utilities for resource allocation: review and empirical analysis. Bmj. 2009 Jul 22;339:b2688.

43 Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. Pharmacoeconomics. 2013 Aug 1;31(8):677-91.

44 Neumann PJ, Stone PW, Chapman RH, Sandberg EA, Bell CM. The quality of reporting in published cost-utility analyses. Ann. Intern. Med. 2000;132(12):964-72.

45 Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Medical Decision Making. 2007 Jul;27(4):491-9.

46 Kassirer JP, Moskowitz AJ, Lau J, Pauker SG. Decision analysis: a progress report. Annals of Internal Medicine. 1987 Feb 1;106(2):275-91.

47 McTaggart-Cowan H, Teckle P, Peacock S. Mapping utilities from cancer-specific health-related quality of life instruments: a review of the literature. Expert review of pharmacoeconomics & outcomes research. 2013 Dec 1;13(6):753-65.

48 Chan KK, Willan AR, Gupta M, Pullenayegum E. Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies. Medical Decision Making. 2014 Oct;34(7):863-72.

49 Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health and quality of life outcomes. 2013 Dec;11(1):151.

50 Sarfaty M, Leshno M, Gordon N, Moore A, Neiman V, Rosenbaum E, Goldstein DA. Cost effectiveness of nivolumab in advanced renal cell carcinoma. European urology. 2018 Apr 1;73(4):628-34.

51 Harvard School of Public Health. The CEA registry: standardizing the methods and practices of cost-effectiveness analysis. Available from URL: https://chds.hsph.harvard.edu/approaches/cost-effectiveness-analysis/ [accessed on May 2019].

52 Kwon JS, Carey MS, Goldie SJ, Kim JJ. Cost-effectiveness analysis of treatment strategies for Stage I and II endometrial cancer. Journal of Obstetrics and Gynaecology Canada. 2007 Feb 1;29(2):131-9.

53 Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Jama. 2002 May 8;287(18):2382-90.

54 Gold MR, Franks P, McCoy KI, Fryback DG. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. Medical care. 1998 Jun 1:778-92.

55 Van Vliet EP, Steyerberg EW, Eijkemans MJ, Kuipers EJ, Siersema PD. Detection of distant metastases in patients with oesophageal or gastric cardia cancer: a diagnostic decision analysis. British journal of cancer. 2007 Oct;97(7):868.

56 Wallace MB, Nietert PJ, Earle C, Krasna MJ, Hawes RH, Hoffman BJ, Reed CE. An analysis of multiple staging management strategies for carcinoma of the esophagus: computed tomography, endoscopic ultrasound, positron emission tomography, and thoracoscopy/laparoscopy. The Annals of thoracic surgery. 2002 Oct 1;74(4):1026-32.

57 Aballéa S, Boler A, Craig A, Wasan H. An economic evaluation of oxaliplatin for the adjuvant treatment of colon cancer in the United Kingdom (UK). European Journal of Cancer. 2007 Jul 1;43(11):1687-93.

58 Torrance GW, Feeny D. Utilities and quality-adjusted life years. International journal of technology assessment in health care. 1989 Oct;5(4):559-75.

59 Torrance GW. Utility approach to measuring health-related quality of life. Journal of chronic diseases. 1987 Jan 1;40(6):593-600.

60 FEENEY D, TORRANCE G, FURLONG W. Chapter 26–Health Utility Index. Quality of life and pharmacoeconomics in clinical trials. Second Edition. Lippincott-Raven Publishers, Philadelphia. 1996.

61 Bristow RE, Santillan A, Salani R, Diaz-Montes TP, Giuntoli II RL, Meisner BC, Armstrong DK, Frick KD. Intraperitoneal cisplatin and paclitaxel versus intravenous carboplatin and paclitaxel chemotherapy for Stage III ovarian cancer: a cost-effectiveness analysis. Gynecologic oncology. 2007 Sep 1;106(3):476-81.

62 Armstrong DK, Bundy B, Wenzel L, Huang HQ, Baergen R, Lele S, Copeland LJ, Walker JL, Burger RA. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. New England Journal of Medicine. 2006 Jan 5;354(1):34-43.

63 Du Bois A, Lück HJ, Meier W, Adams HP, Mobus V, Costa S, Bauknecht T, Richter B, Warm M, Schröder W, Olbricht S. A randomized clinical trial of cisplatin/paclitaxel versus carboplatin/paclitaxel as first-line treatment of ovarian cancer. Journal of the National Cancer Institute. 2003 Sep 3;95(17):1320-9.

64 Durkee BY, Qian Y, Pollom EL, King MT, Dudley SA, Shaffer JL, Chang DT, Gibbs IC, Goldhaber-Fiebert JD, Horst KC. Cost-effectiveness of pertuzumab in human epidermal growth factor receptor 2–positive metastatic breast cancer. Journal of Clinical Oncology. 2016 Mar 20;34(9):902.

65 Casarett D, Fishman J, O'Dwyer PJ, Barg FK, Naylor M, Asch DA. How should we design supportive cancer care? The patient's perspective. Journal of Clinical Oncology. 2008 Mar 10;26(8):1296-301.

66 Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: comparison of assessment methods. Medical Decision Making. 1984 Aug;4(3):315-29.

67 Chuang L, Whitehead SJ. Mapping for economic evaluation. Br Med Bull. 2012;101:1–15.

68 BartonGR,SachTH,JenkinsonC,etal.Doestimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? Health Qual Life Outcomes. 2008;6:51.

69 De Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. International journal of cancer. 1991 Oct 21;49(4):538-44.

APPENDICES

Table S1. Data provided by the CEVR CEA Registry that were used in this study

Item	Description & Coding
Article and Methods Section	
Article ID	ID number in the registry
PubMed ID	
Title of article	
Primary author last name	
Primary author first name	
Journal abbreviation	
Published date	
Primary disease ID (DID)	
Country of study	
Affiliation of author(s)	Classification: academic organizations; health care organizations; governmental organizations; Pharmaceutical, biotech or medical device companies; Contract researcher/consultant; Other; no information. 0=False, 1=True
Study sponsorship/funding	Classification: governmental organizations; foundation/nonprofit organizations; pharmaceutical, biotech or medical device companies; health care organizations; professional membership organization; paper explicitly states no funding; no information; other. 0=False, 1=True
Time horizon	
Time horizon state ID	Stated time horizon clearly. 1=Yes, 2=No
Time horizon units ID	0=lifetime, 1=weeks, 2=months, 3=years
Time horizon magnitude	Number of time unit.
Reader perspective ID	Perspective judged by the registry. 1=Societal, 2=Health care payer, 254=Not stated/could not determine, 250=Other.
Incremental analysis not reported	Incremental analyses were not reported. 0=False, 1=True

Incremental analysis incalculable	Sufficient data were not provided in the study to repeat the ICER calculation. 0=False, 1=True
Sensitivity analysis	Classification: bounded SA (best-and-worst case scenario), PSA, univariate SA (one-way SA), multivariate SA, no SA, unspecified SA. 0=False, 1=True
Utility Weight Section *	
Data source of the weight	Classification: primary data, secondary data, not stated or could not determine.
Direct elicitation	
Direct elicitation ID	Is the weight based on direct elicitation? 0=No, 1=Yes, 3= Not stated or could not determine
Elicitation methods (if based on direct elicitation)	Classification: TTO, Person Trade Off, magnitude estimation, author/clinical judgment, RS, unknown/not stated, not applicable, other.
Measurement scale(s) used	Classification: QWB scale, EQ-5D, SF-36/12/6D, HUI, Generic health status instrument, other generic instrument, other non-generic instrument, unknown/not stated, not applicable.
Sample population type	Classification: community, clinician judgment, author, patient relatives/proxies, unknown/not stated, other.
Sample size	This refers to the utility study sample size.
Weight range	Lower and upper values of weight range.

* Instead of being used directly, utility weight data from the registry was used as reference only, in order to control the internal data validity during data collection.

Table S2. Number of CUAs for cancer/neoplasms included in the review and initially identified in the CEVR CEA Registry*

Pub year	No. of Included CUA	Total No. of CUA in Registry
1988	1	1
1989	0	0
1990	1	2
1991	4	4
1992	2	2
1993	4	4
1994	2	2
1995	2	6
1996	7	7
1997	10	10
1998	7	9
1999	5	7
2000	13	17
2001	10	10
2002	11	12
2003	16	17
2004	17	21
2005	30	31
2006	24	26
2007	40	45
2008	38	42
2009	39	48
2010	58	62
2011	52	59
2012	82	94
2013	102	120
2014	115	125
2015	113	151
2016	107	128
Total	912	1062

* There is no CUA of cancer included in the CEVR CEA Registry from 1976 to 1987.