REVIEW ARTICLE

Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010

**Torbjørn Wisløff · Gunhild Hagen ·
Vida Hamidi · Espen Movik · Marianne Klemp ·
Jan Abel Olsen**

Abstract Reimbursement agencies in several countries now require health outcomes to be measured in terms of quality-adjusted life-years (QALYs), leading to an immense increase in publications reporting QALY gains. However, there is a growing concern that the various 'multi-attribute utility' (MAU) instruments designed to measure the Q in the QALY yield disparate values, implying that results from different instruments are incommensurable. By reviewing cost-utility analyses published in 2010, we aim to contribute to improved knowledge on how QALYs are currently calculated in applied analyses; how transparently QALY measurement is presented; and how large the expected incremental QALY gains are. We searched Embase, MEDLINE and NHS EED for all cost-utility analyses published in 2010. All analyses that had estimated QALYs gained from health interventions were included. Of the 370 studies included in this review, 48 % were pharmacoeconomic evaluations. Active comparators were used in 71 % of studies. The median incremental QALY gain was 0.06, which translates to 3 weeks in best imaginable health. The EQ-5D-3L is the dominant instrument used. However, reporting of how QALY gains are estimated is generally inadequate. In 55 % of the studies there was no reference to which MAU instrument or direct valuation method QALY data came from. The methods used for estimating expected QALY gains are not transparently reported in published papers. Given the wide variation in utility scores that different methodologies may assign to an identical health state, it is important for journal editors to require a more transparent way of reporting the estimation of incremental QALY gains.

T. Wisløff (✉) · G. Hagen · V. Hamidi · E. Movik ·
M. Klemp · J. A. Olsen
Norwegian Knowledge Centre for the Health Services,
Postboks 7004 St. Olavs plass, 0130 Oslo, Norway
e-mail: twisloff@gmail.com

T. Wisløff
Unit of Biostatistics, Epidemiology and Health Economics,
Oslo University Hospital, Oslo, Norway

T. Wisløff
Department of Health Management and Health Economics,
University of Oslo, Oslo, Norway

G. Hagen
Department of Public Health and General Practice, Norwegian
University of Science and Technology, Trondheim, Norway

M. Klemp
Department of Pharmacology, University of Oslo, Oslo, Norway

J. A. Olsen
Institute of Community Medicine, University of Tromsø,
Tromsø, Norway

Key Points for Decision Makers

The incremental quality-adjusted life-year (QALY) gains from medical technologies are generally small

Journal editors should require better transparency in the reporting of how QALY gains have been measured

The EQ-5D-3L is the most widely used instrument for measuring health-related quality of life.

1 Introduction

Over the last 2 decades there has been an immense increase in the number of published cost-utility analyses (CUAs, i.e.

△ Adis
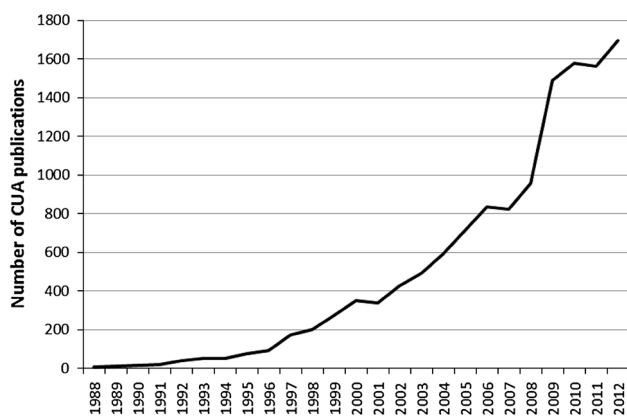
Fig. 1 The growth of cost-utility analyses over the last 25 years (search for cost-utility analysis in Embase)

the type of cost-effectiveness analyses [CEAs] that measure health outcomes in terms of quality-adjusted life-years [QALYs]). To illustrate this, we searched for CUAs in Embase (Fig. 1), resulting in 37 hits from 1992, 425 from 2002 and 1,694 from 2012. This growing interest in expressing outcomes in QALY terms may be explained by a combination of research innovations and policy guidelines from reimbursement agencies [1].

Around 1980, some large-scale research projects started developing multi-attribute utility (MAU) instruments intended to measure health states on a scale of 0–1 [2]. The motivation was to make health gains comparable across symptoms and diagnoses. Much research effort was devoted to constructing both generic descriptive systems and valuation methods to assign values to the 'health-related quality of life' (HRQoL, for short simply Q) associated with being in various health states. The literature refers to these numbers using different terms, such as 'health utility indices', 'quality of life values', 'utility values', 'health state scores' or simply QALY weights.

Principally, six MAU instruments have been developed (EuroQoL-5 Dimensions [EQ-5D], Health Utilities Index [HUI], Short-Form-6 Dimensions [SF6D], Assessment of Quality of Life [AQoL], 15 Dimensions [15D], Quality of Well-Being [QWB]), all based on different descriptive systems and using different valuation methods. One major challenge is evident in the literature that has compared the alternative instruments: they yield different utility scores for the same respondent for the same health state [3–5]. Such differences can be explained by descriptive systems that cover different domains of health, and valuation techniques (visual analogue scale [VAS], standard gamble [SG], time trade-off [TTO], person trade-off [PTO]) that produce different preference scores [6].

When different instruments produce different numbers for the Q in the QALY formula, decision makers are faced with incommensurable analyses. Government bodies in some countries have therefore issued guidelines as to which instrument should be applied in the estimation of QALY gains. For instance in the UK, the National Institute for Health and Care Excellence (NICE) recommends the EQ-5D.

Methodological transparency becomes paramount to make comparisons of study results. Interestingly, in their key text on methods for economic evaluation, Drummond et al. [7] provide a widely used checklist that includes questions on whether outcomes were 'measured accurately' and 'valued credibly'. An important motivation behind the current review is to unravel the extent to which the estimation of QALY gains in published CUAs was reported in a transparent way. More specifically, the aim of this paper is to examine three key questions: (i) How transparent is the reporting of how QALYs are being estimated? (ii) How is the Q in the QALY currently measured in published CUAs? (iii) What is the size of the QALY gain reported?

The threshold value of the incremental cost-effectiveness ratio (ICER), or society's willingness to pay for a QALY, has attracted much interest in the literature [8]. However, beyond the size of the ratio, there are many reasons why the size of the denominator may have policy relevance in its own right. Buyx et al. [9] recently presented the case for a 'minimum effectiveness threshold' of 3 months' additional lifetime. Furthermore, there is an increasing interest in the literature on assigning distributive weights to QALYs, i.e. whether it is better to give large gains to some few or small gains to the many [10]. In other words, do people assign a constant marginal utility to increasing QALY gains [11]? To bring some policy-relevant context into this debate, it is worth knowing the size of QALY gains in published analyses. In this context, it is important to notice that QALY gains reported in CUAs represent the average gain in the specified study population.

2 Methodology

We searched the databases MEDLINE and Embase using medical subject headings (MeSH) and Emtree terms and text words related to economic evaluation and preference-based QALY instruments. More specifically, we used the MeSH and Emtree terms and text words to describe any type of economic evaluation in combination with text words for the six most frequently used MAU instruments (EQ-5D, HUI, SF-6D, AQoL, 15D, QWB [2]) or the four principal valuation methods (VAS, TTO, SG, PTO). For convenience, the search was limited to the year 2010 and thus provides a cross-sectional picture. Realizing that the search may be missing some studies, we also searched the

National Health Service Economic Evaluation Database (NHS EED) for 'cost-utility'.

Two reviewers read the abstracts and excluded studies based on predefined criteria. The key inclusion criteria in the review process were that papers should be published in peer-reviewed English-language journals, and report from an applied CUA, i.e. it should be an original health economic evaluation with costs and QALYs as outcomes. Studies were not excluded based on any data reported (or not reported) in the results section; if an original CUA was performed, the study was included.

Data were extracted from the studies to address the three key questions of our objective:

1. How transparent is the reporting of how the QALY has been measured?
2. How is the Q in the QALY currently measured in published studies?
3. What is the size of the QALY gain in published studies?

More specifically, we explore (i) which MAU instrument was used (EQ-5D-3L, HUI, SF6D, AQoL, 15D, QWB), (ii) on which valuation method (TTO, SG, VAS, PTO) were health utility values based, (iii) what is the time horizon over which the QALY gain is estimated?; and (iv) which discount rates were used to time-adjust future health benefits. Furthermore, we report the variation in the estimated QALY gains, and look more closely into the studies with the highest QALY gains to assess if these were due to important medical breakthroughs.

To address these key issues, we provide some characteristics of the published studies: the main type of disease, country of origin, intervention type and type of journal. These characteristics were hypothesized to possibly influence some of our key questions, such as methodological transparency and size of QALY gain. For instance, it has previously been found that articles published in journals with low impact more often report favourable cost-effectiveness results [12].

Most of the questions we posed could be answered by providing frequency tables. In addition, we analyzed the following differences with simple Chi-square tests:

1. Do health economic journals provide better reporting of which MAU-instrument and valuation technique has been applied?
2. Are large gains more common when the comparator is placebo or no treatment?
3. Are large gains more common in any specific type of journal?

All analyses were performed using SPSS (PASW Statistics 18).
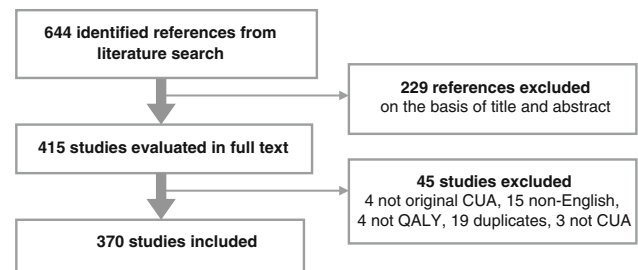
Fig. 2 Search diagram. *CUA* cost-utility analysis, *QALY* quality-adjusted life-year

3 Results

In total, 644 studies were identified. After exclusion of studies that did not meet our inclusion criteria, data were extracted from 370 studies (Fig. 2).

3.1 Characteristics of the Included Publications

Pharmaceuticals or not In total, 176 (47.6 %) studies dealt with pharmacological interventions. The dominance of pharmacoeconomic evaluations was expected, considering that pharmaceutical companies in many countries are obliged to submit a pharmacoeconomic evaluation as part of a reimbursement application, while devices and procedures in many jurisdictions still lack this kind of regulation.

Types of journals The journals in which the included papers had been published were categorized into three main types: (i) clinical or medical specialty journals, (ii) non-specialty medical and health journals, and (iii) health economics type journals (i.e. those with 'economics' or 'technology assessment' in their journal name, as well as *Social Science and Medicine* and *Value in Health*). Only 24 % of the articles were published in this latter type of journal (Table 1).

Country of origin Almost 70 % of the studies had their origin in four countries: the USA (29 %), the UK (23 %), Canada (8 %) and the Netherlands (8 %). The large proportion of studies from the UK, Canada and the Netherlands might be explained by a combination of strong involvement in developing MAU instruments as well as policy guidelines that recommend the use of QALYs in applications for reimbursements. Also, note that the only health technology assessment (HTA) reports identified with our search were from the UK, as these are indexed by MEDLINE and Embase [13]. HTA agencies from other countries may have published some of their analyses through journal articles, but generally HTA reports are disseminated in publications that are not found in searches of regular databases [14].

Type of study Health economic evaluations can be performed as part of an epidemiological study (most often a

Table 1 Descriptives of included studies (*n* = 370)

Descriptive	*n*	%
Intervention		
Pharmaceutical	176	47.6
Not pharmaceutical	194	52.4
Journal type		
Clinical/specialist	209	56.5
General (non-specialist) medical	72	19.5
Health economics type	89	24.1
Country of analysis		
USA	106	28.6
UK	86	23.2
Canada	30	8.1
The Netherlands	29	7.8
Rest of the world	106	28.6
More than one country	13	3.5
Type of study		
Model	297	80.3
Epidemiological study (35 RCTs)	37	10.0
RCT and model	36	9.7
Disease group		
Cancer	71	19.2
Cardiovascular diseases	50	13.5
Respiratory diseases	22	5.9
Mental health	16	4.3
Other chronic diseases	130	35.1
Non-chronic diseases	41	11.1
Lifestyle interventions	18	4.9
Other prevention	22	5.9
Comparator		
An active comparator	218	58.9
Several comparators	44	11.9
Placebo	20	5.4
No treatment	88	23.8

RCT randomized controlled trial

randomized controlled trial [RCT]), as a modelling exercise based on a synthesis of published data, or as a combination of the two. Most studies included in this review were models (80 %). The remainder were split equally between the other two groups: strictly based on an RCT with no modelling involved (10 %) or a combination of RCT and modelling (10 %). The implication of this is that a maximum of 20 % of the evaluations could have had access to individual patient-level QALY data if these were gathered in the RCT, while the rest, in general, would be based on previously published data from one or more sources.

Main disease group The disease groups targeted by the interventions analyzed were categorized in eight groups:

cancer, cardiovascular diseases (CVDs), respiratory diseases, mental health, other chronic diseases, non-chronic diseases, lifestyle interventions and other prevention. Clearly, the vast majority of studies focused on various types of chronic diseases, with cancer and CVD being the most frequent; 19 and 14 % of the studies, respectively.

Comparator With only 5 % placebo controlled and 24 % compared with a no-treatment situation, the vast majority used an active comparator. However, it is impossible to tell if the comparators were chosen to reflect the most relevant alternative.

3.2 Measuring and Valuing Quality-Adjusted Life-Year (QALY) Gains

In 205 studies (55 %), there was no reference to which MAU instrument or direct valuation method formed the basis for measuring 'the Q in the QALY' (Table 2). However, most of these (147) papers had referred to other publications from which they had obtained QALY data. The remaining 58 were based on 'mapping' from a disease-specific instrument, or the valuation method was not specified at all.

Among the studies that explicitly referred to the use of MAU instruments, the EQ-5D-3L was the most frequently used: 87 of the 113 studies that were based on a single instrument. In 11 studies, more than one MAU instrument had been combined. The valuation method used for calculating HRQoL was reported in only 85 (23 %) publications. TTO was the most widely used method, reflecting the fact that most of these studies had applied the standard EQ-5D-3L tariff from the UK, which is TTO based [15].

The combined information on which MAU instrument and which valuation method had been applied was reported in only 66 studies (18 %) (the 16 studies that stated direct valuation and valuation method are included here). Either MAU or valuation method was reported in 99 studies.

The combined information (which MAU instrument and which valuation method) was reported in 29 % of publications in health economics journals, but in only 14 % of medical journal publications (Table 3). Hence, reporting was clearly better in health economics journals (*p* = 0.0013).

3.3 The Size of QALY Gains

In 37 of the 370 studies included, the size of the incremental QALY gain was not reported. Rather, these 37 studies reported the total QALY gain in the study group, the probability that the intervention is cost effective or simply the ICER.

Table 4 shows that the median incremental QALY gain in the remaining 333 studies was 0.06, which translates to 3

Table 2 Multi-attribute utility instrument and valuation method ($n = 370$)

	Valuation method					Total
	TTO	VAS/rating scales	SG	PTO	Missing	
Descriptive instrument						
EQ-5D	31	11	0	0	45	87
SF-6D	0	1	3	0	9	13
HUI	0	0	0	0	6	6
15D	0	1	0	0	4	5
QWB	0	0	0	0	2	2
AQOL	0	0	0	0	0	0
More than one instrument	1	2	0	0	8	11
Mapping	2	3	2	0	35	42
Values from other cited publications	8	0	4	0	147	159
Direct valuation	11	1	3	1	6	22
Not specified	0	0	0	0	23	23
Total	53	19	12	1	285	370

AQOL assessment of Quality of Life, *EQ-5D* EuroQol-5 Dimensions, *HUI* Health Utility Index, *PTO* person trade-off, *QWB* Quality of Well-Being, *SF-6D* Short Form-6 Dimensions, *SG* Standard Gamble, *TTO* time trade-off, *VAS* visual analogue scale, *15D* 15 dimensions

Table 3 Details on utility method reported and journal type

Details on utility method reported	Journal type			Total
	Health economics	Clinical/specialist	General	
MAU and valuation	26 (29)	30 (14)	10 (14)	66 (17)
MAU or valuation	30 (34)	48 (23)	21 (29)	99 (37)
Not reported	33 (37)	131 (63)	41 (57)	205 (46)
Total	89 (100)	209 (100)	72 (100)	370 (100)

Data are presented as *N* (%)

MAU multi-attribute utility

weeks of prolonged life in best imaginable health (the mean was 0.31 QALYs). The effect in the lowest quartile translates to 4 days of prolonged life, while the upper quartile was about 4 months or more. The generally low QALY gains might be due to short time horizons over which the gain had been measured and estimated. Table 4 shows that gains are increasing with time horizon, but not much: the median QALY gain in studies with a time horizon longer than 5 years was only 0.12.

When comparing QALY gains across diagnostic groups (Table 5), we note that interventions related to other chronic diseases yield the highest incremental gains while preventions yield the lowest.

Given the generally low QALY gains in this review, we looked closer into the 29 studies (8 %) that reported a gain larger than 1 QALY. These large gains were most common when the comparator was placebo or no-treatment (14 vs. 7 %, $p = 0.03$). A further characteristic of these studies was that most of them were published in health economic

journals (15 vs. 6 %, $p = 0.01$), which indicates more methodological transparency. Furthermore, a higher proportion of these 29 studies were based on data from 'rest of the world' (14 vs. 7 %, $p = 0.05$), i.e. all countries except for those explicitly mentioned in Table 1.

Eight studies reported incremental gains of two QALYs or more. None of these involved a 'large medical breakthrough'. Rather they were interventions targeted at relatively young patient groups who will benefit from an improved HRQoL over many years. Six of these eight studies compared the gains with a no-treatment alternative.

3.4 Discounting QALY Gains

Discounting QALY gains was common: 276 (75 %) reported results using a positive discount rate, while 58 (17 %) presented only the undiscounted result. In the remaining 36 studies, the discounting issue was not explicitly mentioned.

Table 4 Quality-adjusted life-year gains vs. time horizon

Quartiles of QALY gain	Time horizon levels				
	0.00–1.00 (n = 62)	1.01–5.00 (n = 55)	5.01–Lifetime (n = 190)	Missing (n = 26)	Total (n = 333)
25th percentile	0.01	0.02	0.02	0.01	0.01
Median	0.02	0.08	0.12	0.07	0.06
75th percentile	0.04	0.24	0.47	0.45	0.32
Mean	0.04	0.17	0.43	0.36	0.31

QALY quality-adjusted life-year

Table 5 Quality-adjusted life-year gains vs. disease group

Quartiles of QALY gain	Disease group							
	Cancer (n = 63)	CVD (n = 47)	Respiratory (n = 21)	Mental health (n = 15)	Other chronic diseases (n = 120)	Non-chronic diseases (n = 32)	Lifestyle interventions (n = 16)	Other prevention (n = 19)
25th	0.02	0.01	0.00	0.03	0.02	0.01	0.01	0.00
Median	0.07	0.07	0.04	0.03	0.10	0.04	0.03	0.01
75th	0.33	0.32	0.25	0.20	0.56	0.28	0.15	0.03

CVD cardiovascular disease, *QALY* quality-adjusted life-year

The most frequently applied discount rate was 3.0 %, which has come to be the current standard rate in the literature, perhaps due to recommendations by the Washington panel [16]. Interestingly, studies that departed from this international norm appeared to do so in response to domestic guidelines. The Netherlands suggests a rate of 1.5 % in their guidelines, which explains why 17 of 27 studies using this rate have a Dutch setting. Similarly, the UK guideline is 3.5 %, which explains why 52 of 62 studies using this rate are UK based.

When comparing the practice of discounting with time horizon (Table 6), we note that most studies that did not discount the gains, or contained no mention of discounting (missing), were short-term studies (time horizon of 1 year or less).

4 Discussion

4.1 Transparency

This review of 370 recently published CUAs shows that most QALY calculations are not reported in a sufficiently transparent way. The MAU instruments that were the basis for estimating the Q in the QALY were reported in only one-third of the published studies. In this journal's checklist for modelling studies, the question "Have you detailed the methods that were used to obtain utility values?" is obviously poorly answered, as only 19 % had reported the

Table 6 Discount rates and time horizon levels

Discount rate	Time horizon levels				
	0.00–1.00	1.01–5.00	5.01–Lifetime	Missing	Total
0.0	42 (64)	5 (9)	7 (3)	4 (11)	58 (16)
1.5	1 (2)	9 (16)	16 (8)	1 (3)	27 (7)
3.0	1 (2)	19 (33)	118 (56)	19 (51)	157 (41)
3.5	4 (6)	13 (22)	41 (20)	4 (11)	62 (17)
4.0			1 (0.5)		1 (0.3)
5.0	2 (3)	6 (10)	17 (8)	4 (11)	29 (8)
Missing	16 (24)	6 (10)	9 (4)	5 (14)	36 (9)
Total	58 (100)	50 (100)	175 (100)	37 (100)	370 (100)

Data are presented as *N* (%)

combined information on which MAU instrument and which valuation method were used.

In 43 % of the studies, references were provided for readers and reviewers to search for original sources of QALY weights themselves. We can only speculate as to why such a large share of publications report references, yet fail to follow guidelines that require the reporting of which MAU instrument and which valuation method had been applied. One possible reason for this lack of transparency may be that authors hide facts regarding poor QALY data, i.e. combination of different MAU instruments, or references to other publications that are also insufficient in their methodological reporting. Other reasons may be that authors are unaware of guidelines for publishing economic evaluations, or that they have read

guidelines that are not sufficiently specific regarding the reporting of MAU instrument and valuation method. Recently published guidelines for reporting of health economic evaluations [17] seem to be somewhat more explicit on methodological transparency related to measuring QALYs than the checklist provided in the most widely cited text [7].

We expected that the incremental QALY gain, or at least data from which this could be estimated, would be reported in all publications. However, in 37 studies (10 %), we were not able to find (or calculate) any incremental QALY gain. These studies had either reported the total QALY gain in the study group or the probability that the interventions were cost effective. Clearly, to only report total QALY gain does not comply with guidelines. To only report probability of cost effectiveness is a limitation not explicitly mentioned in most guidelines, but as pointed out by Claxton et al. [18], "the intervention with the highest probability of being cost-effective is not always the one with the highest expected (i.e. mean) cost-effectiveness." Hence, reporting only the probability of being cost effective is insufficient, unless the explicit goal is to prioritize interventions according to probability of being cost effective, rather than according to cost effectiveness [19].

Our review shows that many published CUAs report neither methods nor results as recommended in guidelines. Although our review does not explicitly report on the same breadth of information (or details) as, for example, the Centre for Reviews and Dissemination database [20], our findings suggest that everyone involved in the publication process, i.e. authors, reviewers and editors, should adhere to guidelines more strictly.

4.2 Describing and Valuing Health

Of the six MAU instruments, the high frequency of EQ-5D use is in accordance with what others have found earlier, both in reviews of published literature [21, 22] and in reimbursement submissions [23]. The popularity of the EQ-5D can be explained by many factors, one of which might simply be its practicality. As the shortest instrument, it occupies the least space in questionnaires that are likely to already include a whole range of questions. However, the wide use of an instrument is not necessarily an indicator of quality. In an attempt to improve its quality, the descriptive system has now been refined from three to five levels (EQ-5D-5L).

Ten publications reported that two or more MAU instruments had been used. More studies might also have done so, while reporting the estimated incremental QALY gains from only one instrument. The motivation for using more than one instrument might be to increase the probability of detecting an effect, as expressed in one publication: "Of the two generic, preference-weighted, health-related quality-of-life measures (standard gamble preference-weighted SF-12 and QWB scale), the intervention effect was only significant for the SF-12 QALY and therefore only the SF-12 QALY results are presented." [24]. Interestingly, the incremental QALY gain reported in this particular study was only 0.018 (<1 week).

4.3 The Size of the Incremental Gain

The calculated QALY gains appear to be quite small, which may indicate that large breakthroughs in health science are rare. However, our review of reported QALY gains indicates a wide variation in average expected gain from interventions. When assessing sizes of QALY gains, it is important to stress that these are mean values that do not reveal how gains are distributed across patients. We may have some idea based on the nature of the interventions, i.e. a screening intervention would most likely have a large benefit for the few that have a true positive test result and give no benefit or even a negative benefit for the healthy ones. For some chronic diseases, lifetime might not be lengthened but the quality of life can be importantly improved by new interventions or drugs.

More recently, some suggestions have been made about introducing a minimum level of life extension in new treatments to obtain public funding, or preferential public funding. Based on a small German survey, Buyx et al. [9] present the case for a 'minimum effectiveness threshold' of 3 months additional lifetime. Such a magnitude of gain translates to an incremental QALY gain of 0.25, provided that the increased lifetime involves 'best imaginable health'. In our review, only 30 % of published studies showed gains with at least this magnitude. In the UK, the same magnitude of life extension is required for classification as 'end-of-life treatments', for which higher costs per QALY will be accepted [25]. Here, an additional requirement of a maximum of 2 remaining years of expected lifetime should also be met.

4.4 Limitations of Our Review

Our review examined only CUAs published in 2010. This may not be a representative sample, because some technologies might be over-represented if they were newly introduced in that period or if there was a particular global focus on a given disease.

When performing literature searches, there is most often a trade-off between sensitivity and specificity: if you aim to find all studies, you will have to read through thousands of hits to be certain that none has slipped. In the present article, we decided that specificity would be our priority.

In the articles where only references to sources of QALY data were given, there is a potential for investigating further in order to find more exact data. This is a limitation of our paper with regards to the completeness of the data. However, we believe that to comply with guidelines, and to be appropriately transparent, published papers should report which MAU instrument and which valuation method have been applied.

Given this very large sample of 370 published studies, we had to concentrate on some limited topics: which MAU instruments and valuations techniques were used, whether the estimated gains were incremental to an active comparator, and the practice of discounting to adjust for differential timing. There are certainly several other methodological and normative issues to explore in future reviews, e.g. to what degree utility values are based on the preferences of the general public, patients' experiences, or healthcare personnel. However, this type of information is rarely reported in CUAs, but has been reported in a systematic review of empirical studies [26] and a review of pharmaceutical submissions [23].

5 Conclusions

Our review reveals a generally poor transparency in the reporting of how incremental QALY gains are measured and valued.

The EQ-5D-3L is the most widely used MAU instrument, representing 77 % of those studies that reported which instrument had been used.

The median of the incremental QALY gains reported in 370 CUAs from 2010 was 0.06.

References

1. http://www.ispor.org/PEguidelines/index.asp: ISPOR; [22.12.2013].

2. Richardson J, McKie J, Bariola E. Review and critique of health related multi attribute utility instruments. Centre for Health Economics: Monash University; 2011.

3. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Ann Med. 2001;33(5):358–70.

4. Grieve R, Grishchenko M, Cairns J. SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility. Eur J Health Econ. 2009;10(1):15–23.

5. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim JS. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. Med Dec Making. 2010;30(1):5–15. (PubMed PMID: 19843961. Pubmed Central PMCID: 2812696).

6. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.

7. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.

8. Claxton K, Martin S, Soares MO, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the NICE cost effectiveness threshold. York: University of York; 2013.

9. Buyx AM, Friedrich DR, Schone-Seifert B. Ethics and effectiveness: rationing healthcare by thresholds of minimum effectiveness. BMJ. 2011;342:d54.

10. Gyrd-Hansen D, Kristiansen IS. Preferences for 'life-saving' programmes: small for all or gambling for the prize? Health Econ. 2008;17(6):709–20.

11. Kvamme MK, Gyrd-Hansen D, Olsen JA, Kristiansen IS. Increasing marginal utility of small increases in life-expectancy? Results from a population survey. J Health Econ. 2010;29(4):541–8.

12. Bell CM, Urbach DR, Ray JG, Bayoumi A, Rosen AB, Greenberg D, et al. Bias in published cost effectiveness studies: systematic review. BMJ. 2006;332(7543):699–703 (PubMed PMID: 16495332. Pubmed Central PMCID: 1410902).

13. http://www.nets.nihr.ac.uk/programmes/hta: NIHR; [22.12.2013].

14. Grey matters: A practical search tool for evidence-based medicine. CADTH 2013.

15. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108 (PubMed PMID: 9366889. English).

16. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. Oxford: Oxford University Press; 1996.

17. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement. PharmacoEconomics. 2013;31(5):361-367

18. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005;14(4):339–47.

19. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. J Health Econ. 1999;18(3):341–64.

20. Centre for Reviews and Dissemination databases. http://www.crd.york.ac.uk/NIHR_CRDWEB/HomePage.asp [22.12.2013].

21. Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen OP, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. Int J Technol Assess Health Care. 2006;22(2):235–41 (PubMed PMID: 16571199).

22. Maklin S, Rasanen P, Laitinen R, Kovanen N, Autti-Ramo I, Sintonen H, et al. Quality-adjusted life-years for the estimation of effectiveness of screening programs: a systematic literature review. Int J Technol Assess Health Care. 2012;28(2):145–51.

23. Scuffham PA, Whitty JA, Mitchell A, Viney R. The use of QALY weights for QALY calculations: a review of industry submissions requesting listing on the Australian Pharmaceutical Benefits Scheme 2002-4. PharmacoEconomics. 2008;26(4):297–310.

24. Pyne JM, Fortney JC, Tripathi SP, Maciejewski ML, Edlund MJ, Williams DK. Cost-effectiveness analysis of a rural telemedicine collaborative care intervention for depression. Arch Gen Psychiatry. 2010;67(8):812–21.

25. NICE. Appraising life-extending, end of life treatments 2009 [13.06.2013]. http://www.nice.org.uk/media/E4A/79/Supplementary AdviceTACEoL.pdf.

26. Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. PharmacoEconomics. 2013;31(8):677–91.