
Is Quality-Adjusted Life Years (QALY) terminal? A literature review into QALY's criticisms

Philip Astaire Hines¹

¹Department of International Health, CAPHRI, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.

Corresponding author: Philip Astaire Hines;
Address: 11 Park House Gardens, TW1 2DF, Twickenham, London, UK;
Telephone: 0031628779497; Email: p.ahines@student.maastrichtuniversity.nl

Abstract

Quality-adjusted life years (QALYs) have received a growing amount of criticisms. This paper sought to evaluate the future of QALYs through establishing and assessing common critical themes. Electronic searches of PubMed and Web of Science were conducted. The resulting papers were screened for their common critical themes. A total of 19 relevant studies were found, with six common critical themes: direct and indirect assessments report different disease weightings; patients and the public report different disease weightings; underreporting and lack of standardisation in generation of disease weightings; ageism of QALYs and equity vs. efficiency. The critical themes identified call into questions QALY's validity and reliability. However, they do not completely nullify QALYs as a pragmatic tool for Health Technology Assessments and health economics more broadly. They also highlight areas where QALYs can be improved such as standardisation and incorporating social welfare functions into the QALY construct.

Keywords: health technology assessment, quality-adjusted life years (QALY), social welfare function.

Introduction

With European healthcare systems facing increasing complexity and demand, accurate Health Technology Assessments (HTAs) are being seen as ever more important by healthcare providers (1,2). HTAs use clinical outcomes in tandem with economic analysis to assess health technologies. Clinical outcomes and costs can be combined using a Cost Utility Analysis (CUA). The most widely used CUA quantifies the technology's costs and utility using cost per quality-adjusted life year (QALY) (3). QALY is a function of quality (utility) and quantity of life (life years). It calculates one year lived in best possible health (utility) as one QALY. A single QALY is also equal to two years lived at half utility. In this way, it encompasses both the extra life years which a health technology can provide, as well as the quality of life improvements the technology brings. It is usually expressed in terms of cost per QALY gained. In order to quantify the utilities of different health states, a weighting is assigned to them.

The weighting of health states are calculated from the experiences or preferences of patients and/or the public, respectively (4). There are both direct and indirect methods of weighting health states. The three main direct methods are time trade off (TTO), visual analogue scale (VAS) and standard gambling (SG). TTO involves asking a sample of people how many life years they would trade in order to avoid living with a certain health state: usually a disease or disability (5). The visual analogue scale (VAS) involves participants rating a health state on a scale from 0-100: 0 being the worst imaginable, 100 being optimal health. SG involves asking people what risk of dying they would accept in order to be cured of a disease. Correlations between the direct methods have been shown to be significant but vary depending on the subjects, questions and design of the study (6-8).

There are also many indirect methods to weight health states. They comprise multi-attribute questionnaires completed by participants about their own state of health (9). The most well-known of which being the EQ5D. The EQ5D calculates a utility weighting

based on participants ratings about their mobility, self-care, usual activities, pain/discomfort and anxiety/depression. There are alternatives such as the Health Utility Index (HUI) questionnaire.

Different HTAs calculate QALYs using different direct or indirect methods to weight health states. With each method having its own shortcomings and biases, the validity in the weighting of diseases is controversial (9). QALYs have also been criticised as an objective tool: not accounting for values within healthcare provision (10). As a result of these concerns, there has been growing controversy surrounding the usefulness of QALYs in HTAs. The European Union funded the project European Consortium in Healthcare Outcomes and Cost-Benefit Research (ECHO-UTCOMES) to evaluate current HTA practices in Europe. They investigated the underlying assumptions of QALYs as the most prevalent HTA tool in Europe, and recommended their abandonment due to lack of validity (11). In the light of such criticisms, this study will seek to review the literature criticising QALYs in an attempt to establish and assess common critical themes and their implications on QALY's future.

Methods

Data collection

A literature review was carried out using PubMed and Web of Science. The search terms "QALY" OR "quality of life adjusted years" AND "criticism" were used for both. Only research published after 1990 was included, as although QALYs have a fixed conceptual nature, the criticisms had more empirical evidence after this date.

Data selection

The resulting studies were narrowed first by assessing their relevance through the title, then by examining the abstract and finally the content. The inclusion criteria were as follows:

- Contained a reasoned criticism of QALYs and/or a specific example of a failure with QALYs;
- English language;
- Available in full text.

Studies which failed to meet these criteria were excluded. Relevant studies which had previously been read for background research were included.

Data analysis

The criticisms/failures of QALYs were identified in each paper and then amalgamated into common

critical themes.

Results

The literature review combined with background reading produced 22 relevant research papers. Of these, 19 met the inclusion criteria and are listed in Table 1.

Table 1. Critical themes within the literature evaluating QALYs

Critical Theme	Research Papers
<i>Direct and indirect assessments report different disease weightings</i>	Badia, 1999 (7), Craig, 2009 (8), Iskedjian <i>et al.</i> , 2013 (6), Marra <i>et al.</i> , 2004 (9), Ariza-Ariza, 2006 (12), Johnsen, 2013 (13), Fryback <i>et al.</i> , 2008 (14), Beresniak <i>et al.</i> , 2012 (11),
<i>Patients and the public report different disease weightings</i>	Smith <i>et al.</i> , 2008 (15), Drummond <i>et al.</i> , 2009 (16), Kahneman, 2006 (3), Dolan, 2008 (10)
<i>Underreporting and lack of standardisation in generation of disease weightings</i>	Wisloff <i>et al.</i> , 2014 (17), Dolan, 2008 (10), Drummond <i>et al.</i> , 2009 (16)
<i>Ageism of QALYs</i>	Kappel and Sandoe, 1992 (18), Williams, 1997 (19)
<i>Equity vs. efficiency</i>	Wagstaff, 1991 (20), Østerdal, 2005 (21), Harris, 2005 (22), Edlin <i>et al.</i> , 2013 (23)

The results from the data extraction and analysis are presented in Table 1. There were five critical themes which arose more than once within the included papers. The critical theme with the largest representation in the literature criticised the difference in disease weightings given by direct and indirect methods. The next most common theme was QALY's lack of trade-off between efficiency and equity, with four papers focussing on this. The difference between patient (experienced) and public (imagined) disease weightings in QALYs were included in three papers. Three further

papers criticised the lack of standardisation and/or underreporting in the generation of disease weightings. A further two papers discussed the ageism within QALYs. Dolan, 2008 (10) and Drummond, *et al.* 2009 (16) discussed patient vs. public weightings, as well as underreporting and lack of standardisation within the weightings.

Discussion

Below, the themes are evaluated, with reference to any paper which served to forward the criticisms.

Direct and indirect assessments report different disease weightings

People and populations value health states differently, and so within samples undertaking direct and indirect QALY assessments, there will be a lot of heterogeneity. Similarly across different assessment tools such as TTO and VAS, several authors found there to be differences between how people weight diseases (6-9,12-14). This would be less problematic if the results between the assessments correlated and were scalable. Accordingly Fryback *et al.* (14) looked at correlations between most of the major indirect methods to see whether a common scale could be produced. He found that it would only be possible for specific instances when certain levels of health care were being assessed, otherwise the correlations were too modest. It is this variance in disease weightings that led the ECHOUTCOME group to recommend the termination of QALYs in HTAs (11).

However, it is debatable whether the variance in current assessment techniques means that no future tool can be designed to reduce this. Still, it would seem that individual and cultural variance would persist. Dependent on one's perspective, it may also be argued that trying to account for individual weightings is beyond the remit of QALYs as an efficient, objective tool for HTA's.

Patients and the public report different disease weightings

Kahneman (3) discussed QALYs from a behavioural economist standpoint. He raised the point that most QALY's use direct disease weighting, therefore it is the public who are valuing the 'sale' of their health. This is opposed to if it were the patients valuing the 'purchase' of health. Therefore QALY's are subject to the endowment effect, whereby sellers tend to value goods higher than if they were buying those goods. Kahneman (3) also concluded that due to inexperience of the health states, the public will also be subject to the focussing illusion. Meaning people may imagine the health state in its most severe form, prior to adaption: further overweighting the health state.

Dolan (10) echoed this point stating that "...it is much better to ration health care according to real experiences rather than according to hypothetical preferences". However Smith *et al.* (15) contend that subjective (patient) based responses may not be more accurate than direct methods of reporting. They suggest that similar focussing bias will occur when imagining life without the disease. Patients are also open to memory bias: varied interpretations of how the disease state is or was. It would appear that large scale research would need to be conducted, looking directly at the differences in patient and public weighting. With a combination of weightings, from the two being the most representative, although not necessarily accurate weighting.

Underreporting and lack of standardisation in generation of disease weightings

There was concern expressed by Drummond *et al.* (16) and Wisloff *et al.* (17) about the lack of methodological standardisation within disease weighting generation. Drummond *et al.* (16) discussed the need for a 'reference case' which enabled more accurate comparability of QALYs across diseases. However from the above criticisms it would seem that none of the current assessment tools would be suitable for this. In a literature review by Wisloff, Hagen (17) of the 370 studies looking at producing QALY values, 55% of studies did not include which assessment tool was used. They recommended that journal editors require transparency on methods used to weight health states. Standardisation of QALY generation may serve to increase its reliability as a comparative tool. Furthermore standardisation of assessment techniques may also improve transparency around QALY generation. Nevertheless the broad nature of diseases and treatments that QALYs encompass will make standardisation incredibly challenging, especially with regards to assuring validity.

Ageism of QALYs

QALYs are sometimes perceived to have inherent

ageism. For example a lifesaving treatment will be worth more QALYs to a person of 20 than someone of 80 as the person aged 20 has many more life years to live. Therefore healthcare providers may be willing to pay for a treatment for the 20 year old, but not the 80 year old. Nonetheless it is not structurally ageist, as a QALY to someone of 80 is equal to someone of 20. Kappel and Sandoe (18) and Williams (19) call for an ageist bias towards the young to be actively built into QALY generation. Kappel and Sandoe (18) argue on consequentialist grounds that resources are more productive if given to the young. This seems a valid argument and other instruments such as disease adjusted life years incorporate different age weightings into their assessments. Nonetheless, this may be politically untenable in an environment where QALYs are already perceived to be bias against the old. However Williams (19) call for an age bias towards the young on the grounds of the 'fair innings principle': the notion that people are entitled to a certain amount of life years. This concept may be a more socially acceptable explanation for incorporating an age bias into QALY.

Equity vs. efficiency

QALYs are structured to maximise efficiency and are blind to equity issues such as distribution in health care. For example the public may feel that priority in treatment should be given to those in worse health. Yet a 0.1 of a QALY is worth the same for someone subsisting in a health state of 0.2 as someone at 0.8, which may seem perverse. Harris (22) claims that QALYs should be ignored in life saving treatments. He believes that no one should be denied such treatment based on their capacity to benefit. However Edlin, McCabe (23) point out that this ignores the opportunity cost of using those resources on healthcare that will perhaps save more lives in the future, such as breast cancer screening. Wagstaff (20) suggests incorporating a social welfare function (SWF) into the QALY to weight both efficiency and equity. This is developed further by

Østerdal (21) who believes direct methods of health state assessment should include questionnaires about distributive justice to get an idea of a SWF for health. This would appear to be a valid method of balancing equity and efficiency. Incorporating distributional justice questionnaires alongside QALY assessment tools would also be a relatively cost effective method of measuring societal preferences. However such preferences would also be subject to similar biases as seen in the health care assessments.

Strengths and limitations

This review looked only at the criticisms of QALYs, and so provided a viewpoint skewed towards its weaknesses. The rationale behind this being that if the criticisms were defensible then the positives become irrelevant when looking at QALY's future, it also ensured a robust critical review of the QALYs.

The review only included a limited number of papers. There were also more points of criticism that individual papers raised, but which did not constitute a theme on their own. It also did not take into account the opinion of HTA professionals or policy makers who are the actors which will ultimately decide the fate of QALYs (15).

Conclusions

There are many criticisms facing QALYs. Such criticisms undoubtedly weaken the underlying assumptions QALYs are constructed from. However a tool designed to objectively quantify subjective individual preferences and experiences will never be completely sound. Whilst important, these criticisms do not reduce QALY's function completely: it remains a useful, pragmatic tool for valuing treatments within HTAs. They do however highlight areas where QALYs can be improved such as standardisation and incorporating SWFs into the QALY construct. Such improvements would further QALY as a tool and begin to allow valid comparisons of HTAs across healthcare systems. This research would be timely with the

European Union exploring increased cooperation in HTAs (24).

The future of QALYs is not only a function of their reliability and validity, but also an artefact of the limited alternatives we currently have. In a world of

increasingly complex health care systems and treatment options QALYs stand out as useful tool for making comparisons across treatments and conditions. The death of QALY has a “why?”, but a “how?” and “when?” remain to be seen.

Conflicts of interest: None declared.

References

1. Kirschner N, Stephen GP, Stubbs JW. Information on Cost-Effectiveness: An Essential Product of a National Comparative Effectiveness Program. *Ann Intern Med* 2008;148:956-61.
2. Wendt C, Kohl J. Translating Monetary Inputs into Health Care Provision: A Comparative Analysis of the Impact of Different Modes of Public Policy. *JCPA* 2010;12:11-31.
3. Kahneman D. Determinants of health economic decisions in actual practice: the role of behavioral economics. In: Kahneman D, editor. *ISPOR 10th Annual International First Plenary Session*; Washington DC, USA: Value Health; 2006. p. 65-7.
4. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health* 2009;12:27-30.
5. Burström K, Johnson M, Diderichsen F. A comparison of individual and social time trade-off values for health states in the general population. *Health Policy* 2006;76:359-70.
6. Iskedjian M, Navarro V, Farah B, Berbari J, Walker JH, Le Lorier J. PRM134 - Examination of patient preferences: subgroup comparisons of time trade-off and standard gamble results. *Value Health* 2013;16:A37.
7. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Qual Life Res* 1999;8:303-10.
8. Craig BM, Busschbach JJ, Salomon JA. Modeling Ranking, Time Trade-Off, and Visual Analog Scale Values for EQ-5D Health States: A Review and Comparison of Methods. *Med Care* 2009;47:634-41.
9. Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care* 2004;42:1125-31.
10. Dolan P. Developing methods that really do value the ‘Q’ in the QALY. *Health Econ Policy Law* 2008;3:66-9.
11. Beresniak A, Auray J, Duru G, Medina-Lara A, Tarricone R, Sambuc R, et al. PRM14 European Assessment of the Validity of the QALY Outcome Measure: Results From the Experiment Conducted by the Echoutcome Project. *Value Health* 2012;15:A462.
12. Ariza Ariza R, Hernández Cruz B, Carmona L, Ruiz Montesinos DM, Ballina J, Navarro Sarabia, F. Assessing utility values in rheumatoid arthritis: A comparison between time trade off and the EuroQol. *Arthritis Care Res* 2006;55:751-6.
13. Johnsen LG, Hellum C, Nygaard OP, Storheim K, Brox JJ, Rossvoll I, et al. Comparison of the SF6D, the EQ5D, and the Oswestry disability index in patients with chronic low back pain and degenerative disc disease. *BMC Musculoskelet Disord* 2013;14:148. DOI: 10.1186/1471-2474-14-148.
14. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim J-S, editors. *Cross-walks among five self-reported summary health utility indexes: progress and prospects. The 29th Annual Meeting of the Society for Medical Decision Making*; 2008.
15. Smith DM, Brown SL, Ubel PA. Are subjective well-being measures any better than decision utility measures? *Health Econ Policy Law* 2008;3:85-91.
16. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E, Consensus Development Group. *Toward a Consensus on the QALY. Value Health* 2009;12:S31-S5.
17. Wisloff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics* 2014;32:367-75.
18. Kappel K, Sandoe P. QALYs, age and fairness. *Bioethics* 1992;6:297-316.
19. Williams A. Intergenerational equity: an exploration of the ‘fair innings’ argument. *Health Econ* 1997;6:117-32.
20. Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ* 1991;10:21-41.
21. Østerdal LP. Axioms for health care resource allocation. *J Health Econ* 2005;24:679-702.
22. Harris J. It's not NICE to discriminate. *Journal Med Ethics* 2005;31:373-5.

23. Edlin R, McCabe C, Round J, Wright J, Claxton K, Sculpher M, Cookson R. Understanding Harris' understanding of CEA: is cost effective resource allocation undone? *J Health Serv Res Policy* 2013;18:34-9.
24. ECORYS. European Cooperation on Health Technology Assessment Economic and governance analysis of the establishment of a permanent secretariat. Rotterdam: Executive Agency for Health and Consumers, 2013.