

Toward a Consensus on the QALY

Michael Drummond, DPhil,¹ Diana Brixner, BSPHarm, PhD,² Marthe Gold, MD, MPH,³ Paul Kind,⁴
Alistair McGuire, PhD,⁵ Erik Nord, PhD⁶ (Consensus Development Group)

¹University of York, York, UK; ²University of Utah, Salt Lake City, Utah, USA; ³City University of New York Medical School, New York, NY, USA; ⁴University of York, York, UK; ⁵London School of Economics, London, UK; ⁶Norwegian Institute of Public Health, Oslo, Norway

Keywords: resource allocation, quality of life, health state valuation.

Introduction

Earlier versions of the other articles in this *Value in Health* Special Issue, *Moving the QALY Forward: Building a Pragmatic Road* (articles 2–6) were presented at an ISPOR Development Workshop on “Moving the QALY Forward: Building a Pragmatic Road” held in Philadelphia from November 6–8, 2007. After the discussion of the articles, a *workshop consensus group* was formed, including representation from the groups producing the other articles. It fell to the consensus group to identify common ground on key issues.

Given the diversity of views expressed at this workshop, as evidenced in the articles presented, it was clear that it would not be possible to reach agreement on specifics, such as how QALYs should be measured, or which instrument(s) should be used. Rather, the group felt that the best way forward was to reach agreement on several high-level principles and to express disagreements as a set of issues for further research. In all, eight general statements were agreed upon, which were then put to all the workshop participants for discussion. These statements are outlined as follows, along with the main points arising in the discussion.

Consensus Statements

QALYs Are One Health-Based Input to Health and Health-Care Decisions

This statement embodies two important points. First, QALYs, as currently constituted, focus on health as opposed to well-being more generally. This is the most appropriate focus, given that the stated objective of most health-care systems is to improve the health of the population.

Second, the statement recognizes that there are other inputs to health decisions, which need to be considered alongside QALYs. These were not specified in detail, but the existing literature shows that equity and social justice are salient considerations in the allocation of health-care resources. In the discussion, it was pointed out that further research was required into the nature of such considerations and how they should be incorporated (see Statement 5).

Address correspondence to: Michael Drummond, University of York, Centre for Health Economics, Heslington, York, YO10 5DD, UK. E-mail: md18@york.ac.uk

10.1111/j.1524-4733.2009.00522.x

Michael Drummond, Diana Brixner, Marthe Gold, Paul Kind, Alistair McGuire, and Erik Nord have no conflicts to declare.

QALYs Can Be Used at Various Levels in the Health-Care System

It became apparent during the conference that participants felt that QALYs could potentially inform various types of decisions. Traditionally, they have been used to inform broad resource allocation decisions among groups in the population. On the other hand, some participants felt they could be used to inform the choice of treatment for individual patients or patient groups. Within the context of a private health plan, they might be used to inform decisions about coverage or the levels of co-payments.

It is important to recognize these various potential uses of QALYs because this is one source of the disagreement over *whose values* should be used in constructing QALYs. For example, the fact that QALYs are often used for population-wide decisions is usually cited as a reason for eliciting health state preference values from members of the general population. On the other hand, those whose focus is primarily on clinical decisions, involving different treatments for the same condition, tend to argue that the most appropriate values are those of patients.

In the discussion, it was pointed out that although there was some knowledge in the use of QALYs in broad resource allocation decisions (e.g. National Institute for Health and Clinical Excellence (NICE) in the UK), little was known about their use in other decision-making contexts. Therefore, the role of QALYs as a “multilevel” decision tool should be a priority for research, perhaps by undertaking case studies and surveys in different settings. Conference participants noted that in countries where the government was not the primary funder of health care, resource allocation decisions occurred in narrower contexts and there is a need to determine how decision-makers in these settings would make best use of QALYs.

Health Is a Determinant of Well-Being

Although the current focus of the QALY, and most health-care decision-making, is on improvements in health, this statement acknowledges that the broader objective is to improve well-being. Although there was not a consensus on the need to widen the concept of QALYs to encompass all aspects of well-being, there was recognition of the need to think more expansively about the impact of health-care interventions and their relationship to other social programs. This is particularly important in areas such as broad public health interventions, mental health, and care of the elderly.

In the discussion, it was pointed out that despite statements about the need for intersectoral initiatives by governments, relatively little is known about the relationship between health and well-being. Also, little is known about the nonhealth effects of

health-care interventions, or the health impacts of programs in social care, education, or environmental protection.

If the aim was to optimize the use of all resources in public programs, an expanded QALY-type measure of well-being would clearly be useful. However, this remains a matter for further research and should not lead us to neglect the use of the QALY as an important component of current population-based health-care decision-making.

Both Ex Ante Preferences and Experiences May Count

One of the main discussion points during the workshop was whether individuals who had not experienced particular health states could adequately conceptualize them (this discussion relates back to the Issue Panel debate held at the ISPOR 11th Annual International Meeting, May 2006).

This issue cannot be separated from those discussed under Consensus Statement 2, in that members of the general public are unlikely to have experienced all the health states one might ask them to value, although they may have friends or family that have experienced the states concerned. Therefore, it is possible that some researchers may take the *normative* view that, for broad resource allocation decisions, the general population's values should count, while at the same time agreeing with the *positive* statement that those who have experienced particular health states can better conceptualize them.

The consensus group noted that there were several unresolved issues in the measurement of experienced utility. For example, in studies of utility in patients and disabled people, many subjects report unwillingness to sacrifice any life expectancy to become well, even though they are experiencing significant health problems. Therefore, although information about these experiences is important in health-care decision-making, it is not clear how such information could be expressed in terms of scores on a 0 to 1 utility scale [1].

Nevertheless, this consensus statement recognizes that both ex ante preferences (held by those who have never experienced a given health state), as well as the preferences of those that have experienced it, could be relevant, depending on the decision-making context and the normative position taken. In some situations, the two sets of preferences could be combined, in that members of the general public could be informed about the views of those who have experienced the health states that are the subject of valuation.

The consensus group did not discuss how this might be operationalized, but did note that more research is required into individuals' evaluations of their experiences in given health states, how these relate to community preferences, and how they can be better used to inform community preferences.

Distributive Issues Need to Be Addressed

A cost-per-QALY ratio indicates the cost-effectiveness of an intervention. As such, the ratio is a measure of efficiency, rather than of "fairness." There was extensive discussion at the workshop of the challenges raised by the failure of QALYs to account for distributive concerns (e.g., the relative priority given to individuals of different levels of current health, and/or different capacity to benefit in terms of life expectancy or health-related quality of life.) In decisions about resource allocation across patient groups, concerns for fairness will count alongside concerns for efficiency in the production of health. Such concerns may cause social resource allocation preferences to deviate considerably from the ranking that consideration of costs per QALY would suggest. These concerns have been

raised internationally in all settings where cost-effectiveness analysis has been studied and/or applied.

The consensus group recognized that there were at least three ways of addressing distributive issues in using QALYs. First, the method of QALY estimation may include a distributional component in its construction. For example, valuation based on *person trade-off* differs from other valuation methods in that it explicitly addresses choices between groups of individuals suffering from conditions of differing severity. Both theoretical and empirical work has illustrated the manner in which distributional issues can be factored into QALYs using the person trade-off (PTO) and other methods.

Second, QALYs calculated from valuation systems such as the visual analog scale (VAS), the time trade-off (TTO), or the standard gamble (SG) could be weighted in accordance with public values for distributive justice rather than being equally weighted as they are at present. The process for generating weights could be based on research into the preferences of members of the public, or developed through the use of citizens' juries. This work would involve research into which normative factors (e.g., condition severity, capacity to benefit, fair innings) should gain highest priority in creating distributional weights.

Third, QALYs may be retained as a measure of health production and address distributive issues through a deliberative decision-making process that would accompany the process of using QALYs for making determinations of resource allocation. For example, committees deciding on the allocation of health-care resources could be asked to address given equity objectives or constraints, as well as considering the incremental cost-effectiveness ratios of given interventions.

Although the consensus group did not come to a view on the preferred approach, it did identify some of the key considerations in the choice. On the one hand, there is a need for transparency. The more complicated one makes a QALY measure by incorporating a sophisticated system of QALY weights, the less transparent it may be. If one of the current obstacles to the use of QALYs is that decision-makers do not understand them, more complicated formulations might worsen this situation.

On the other hand, the power of quantification may mean that distributional issues are more likely to be taken into account if they feature, in some way, in the overall calculation of the value of a health-care intervention. If these issues are merely noted as something else for the decision-maker to consider alongside the calculation of value, they may not be adequately considered.

There is a need for research that compares the different approaches for incorporating distributional issues into the decision-making process. For example, are decision-makers' ex ante views about equity better accommodated by a QALY estimation procedure based on the PTO? Are distributive issues more likely to be taken into account if they feature in the actual quantification of the total QALYs gained (through an explicit weighting scheme), as opposed to the accompanying decision-making process? How feasible is it to develop an appropriate weighting scheme?

There was broad agreement at the workshop that distributional justice is central to resource allocation, but given the lack of empirical data about the approach that would best serve decision-makers, there was no consensus as to which of these approaches are superior. At present, the status quo leaves distributive concerns outside of cost-effectiveness ratios and many workshop participants regarded maintaining interpretability of QALYs as a priority. However, there is a vital need to engage the decision-making community more effectively in determining how they would like additional distributive information to be incorporated into economic evaluations.

Different Methods for Valuing Health Yield Different Results and This Needs to Be Better Understood

The various methods for eliciting preferences for health states embody differences in: 1) health state descriptive systems; and 2) valuation approach (e.g. TTO, SG, VAS, PTO). The process of describing health is an important prerequisite to establishing explicit values. Although there are many similarities between the various descriptive systems, there are also important differences (for example the omission of social functioning from the Health Utilities Index (HUI) and of energy/fatigue from the EQ-5D).

More research is required on the relationship between the different approaches and their validity. In addition, there is a need to understand the extent to which the different descriptive systems and associated value systems are stable over time and across settings. Also, the translations between the different descriptive systems need to be explored (see later discussion).

The aim of QALYs is to value interventions and thus gains (or averted losses) in health. In the conventional QALY approach, this is done in an indirect manner: one uses healthy people's *ex ante* valuations of health states and takes the difference between the pre- and the post-intervention value as the value of the gain (the "subtraction method"). This approach may fail to capture how individuals value treatment given that they have an illness or value being sure of receiving treatment in case of future illness. The valuation of treatment may be less than proportional to the size of the health effect [1]. Research is required on the issue. Somewhat similarly, one may question the assumption (implicit in QALY calculations) of proportionality between the number of life years gained and value. In this area, as elsewhere, there may be diminishing marginal utility of goods (here: years).

Health Gains Need to Be Aggregated over Time

The consensus group was aware that the traditional approach for aggregating health gains over time is rather simplistic; namely, the value of health states is multiplied by the time spent in each state. More research is needed on the linearity of preferences over time and the ways of obtaining the value of pathways or profiles. The group was aware that, in the debate about healthy years equivalents (HYES) in the 1980s, it was acknowledged that HYES represented a better approach in principle, but was largely unworkable, particularly in the context of state transition models for health-care interventions. Therefore, the question is whether superior, but workable, methods for aggregating health gains over time can be developed.

A "Reference Method" Is Required for QALYs

The background to this statement was that the consensus group felt that the QALY had not reached its full potential, certainly in terms of its adoption by decision-makers. However, given the current level of disagreement about the preferred approach, the group considered that a QALY "reference method" that would inform the "reference case" concept, first developed by the US Panel on Cost-Effectiveness in Health and Medicine [2] is needed if cost-effectiveness analyses were to be made more comparable across diseases and interventions. Under the reference case approach, a standardized approach for estimating QALYs would be proposed for inclusion in all economic evaluations, without excluding other approaches. This would facilitate comparability in economic evaluations without stifling further methodological development in the estimation of QALYs.

The consensus group felt quite strongly that this step was needed in order to move the debate on. Otherwise, it would be

likely that workshops like this one would still be taking place 20 years from now. However, the group was not prescriptive about how a reference method would be arrived at, noting that 1) one could pick a measure, or profile of measures, and adopt it; 2) one could develop a "crosswalk" procedure (or translation) between the most common measures, so that in principle any could be used [3]; or 3) one could establish a group, consisting of both methodologists and decision-makers to review the options and to decide on the best approach.

The consensus group recognized that there were potential methodological and practical risks in proposing a reference method. For example, all of the existing measures (e.g. the EQ-5D, HUI) focus on specific attributes of health/ill health. Therefore, if a single measure were chosen, it could bias the broad allocation of health-care resources in favor of those health-care interventions that have the largest impact on those particular attributes. This bias could be minimized by proposing a reference approach that was based on more than one measure, but this would be more costly (e.g., in collecting data in primary studies such as clinical trials). However, despite recognizing these risks, the consensus group felt that there was no alternative if the field were to move forward.

As one would expect, this consensus statement generated a good deal of discussion. A few workshop participants were not convinced that the selection of a reference method was necessary in order to move the field forward. Others felt that, if a reference approach were to be determined, one should not necessarily be constrained by having to select from the existing measures, all of which had their drawbacks. Some felt that it would be prudent to wait for the results of the existing cross-walk exercises [3] before taking the next step.

Summary

The consensus statements previously mentioned represent agreement on a number of "high-level" points. Most were accepted by the majority of participants, although there were a number of qualifications. The greatest disagreement was over the last statement, where a few participants felt that they could not sign up to a process that would generate a reference method for estimating QALYs at this point in time. There was, however, a substantial majority in favor of taking this step, which was seen as advancing the use of QALYs in cost-effectiveness analyses.

The discussions in the consensus group and the subsequent discussions among all workshop participants generated a list of research questions. These are summarized in Table 1.

Discussion: Identifying Ways Forward

Having responded to the consensus statements and generated the list of research questions, the discussion then turned to some of the practical issues that would need to be resolved if we were to move forward with the QALY. Many of these revolved around current and future health care decision-making on both sides of the Atlantic. Others related to the operationalization of a "reference method" for estimating QALYs.

Current Use of QALYs

It was felt that there should be more study of those organizations, including NICE in the UK, that currently use QALYs [4]. For example, is there any evidence that this leads to "better" decisions? Do those currently using QALYs think they are adequate for the task? What else is considered, in decision-making, alongside the incremental cost-effectiveness ratio (ICER)? What is the

Table 1 Research Priorities for the Future of the QALY

1.	The relevance of nonhealth objectives to health-care decision-making
2.	Case studies on the use of QALYs at different levels in the health-care system
3.	Case studies on the use of QALYs in decentralized, privately funded health-care systems, such as that existing in the US
4.	The impact of health on broader well-being
5.	The role of an expanded QALY incorporating dimensions other than health
6.	The relationship between the valuations of those experiencing, or having had experience of, health states and the valuations of the general public for the same states.
7.	Methods for briefing members of the public on the experiences of those in particular health states
8.	Comparisons between the main methods for valuing health in respect of their incorporation of distributional concerns
9.	Weighting schemes reflecting distributional concerns
10.	Qualitative research into the community's views about distributional issues
11.	Evaluation of health gains vs. the evaluation of health states
12.	Research into the assumption of linearity of preferences over time and the ways of obtaining valuations of pathways or profiles
13.	Development of a reference case, or series of reference cases, for estimating QALYs

experience with citizens' juries in supporting the decision-making process, such as NICE's Citizens Council?

Challenges of Using QALYs in Some Settings

It was noted that most of the use of QALYs had been in centralized, single-payer, health-care systems like the British National Health Service. Several participants were unsure about the role of QALYs in decentralized health-care systems like that existing in the United States, where the main motivation may not be to maximize the health of the total population (subject to equity constraints).

However, other participants pointed out that major payers (e.g., WellPoint), and large managed care organizations (e.g., Kaiser-Permanente) cover a "community" of individuals that is the size of many European countries. Some of these organizations had already considered the impacts of therapy on quality of life in their decision-making processes [5,6]. It was also noted that, although there has been a resistance to considering cost or cost-effectiveness in the past, the mood in the United States was changing, given the growing unaffordability of health care in both public and private sectors. A readiness to consider limitations in health care has been found both with members of the public [7] and within decision-making circles [8,9]. It is also evident in the current discussions about a new US federal initiative in comparative effectiveness research that some parties feel should include studies of cost-effectiveness [10].

The workshop participants felt that the way forward was: 1) to explore with decision-makers their interests in and requirements for making use of QALYs and economic analyses; 2) educate them with respect to advantages and limitations of the method; and 3) conduct case studies in a range of decision-making settings that would inform methods decisions. The decision-making matrix reported in an earlier article in this *Value in Health* Special Issue would be a useful starting point for identifying situations where case studies could be conducted.

Developing a Reference Method of Estimating QALYs

Because the majority of workshop participants were in favor of this, there was some discussion of how best to proceed. First, under the aegis of which group should this development take place? Should it be an international effort, led by an organization

such as ISPOR, or would it need to consist of a series of regional or local efforts, given different decision-makers' needs and the different traditions (comparing the United States and Europe) in the use of particular instruments? Proposed US-based entities that could lead this effort include: the Agency for Healthcare Research and Quality, the Institute of Medicine, and the National Institutes of Health. There was no resolution of this issue.

Secondly, what should be the composition of the group deciding on the reference approach? The general opinion was that it should be broadly based, including economists, epidemiologists and decision-makers representing public health, managed care, employers, and government.

Finally, what should be the time frame? Several participants felt that if the need for more data was used as an excuse for not developing a reference method, it may never happen. The consensus was that movement should go forth quickly with the commitment that any decisions made would be reviewed after a few years.

Conclusions

The workshop was successful in two respects. First, a broad range of perspectives on the future of the QALY were discussed. Secondly, agreement was reached on several high-level principles and a research agenda was outlined.

Nevertheless, it was not possible to agree on a single way forward and, given the diversity of opinion, this is hardly surprising. Nevertheless, the majority opinion was that if the QALY was to gain broader acceptance by decision-makers, something must be done to move the debate forward. To this end, the majority opinion was that efforts should be made to develop a reference method for estimating QALYs. If this does eventually happen, this workshop will have made a major contribution. Only time will tell.

Source of financial support: Funding for the ISPOR "Building a Pragmatic Road: Moving the QALY Forward" Consensus Development Workshop was made possible in part by grant 1R13 HS016841-01 from the Agency for Healthcare Research and Quality. The views expressed in written conference materials or publications and by speakers and moderators neither necessarily reflect the official policies of the Department of Health and Human Services; nor does mention of trade names, commercial practices, or organizations imply endorsement by the US government. Funding for this *Value in Health* Special Issue, "Moving the QALY Forward: Building a Pragmatic Road" was made possible in part by Contract No. HHSN261200800148P from the National Cancer Institute.

Michael Drummond, Diana Brixner, Marthe Gold, Paul Kind, Alistair McGuire, and Eric Nord have no conflicts to declare.

References

- 1 Nord E, Daniels N, Kamlet M. QALYs: some challenges. *Value Health* 2008;12(Suppl. 1):S10–5.
- 2 Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- 3 Fryback DG, Palta M, Cherepov D, et al. Cross-walks among five self-reported summary health utility indexes: progress and prospects. *Med Decis Making* 2008;28:E24.
- 4 Rawlins MD, Culyer AJ. National institute for clinical excellence and its value judgments. *BMJ* 2004;329:224–7.
- 5 Aspinall SL, Good CB, Glassman PA, Valentino MA. The evolving use of cost-effectiveness analysis in formulary management within the Department of Veteran Affairs. *Med Care* 2005;43(Suppl. 7):S20–6.

- 6 Wu WK, Sause RB, Zacker C. Use of health-related quality of life information in managed care formulary decision-making. *Res Social Adm Pharm* 2005;1:579-98.
- 7 Gold MR, Sofaer S, Siegelberg T. Medicare and CEA; time to ask the taxpayers. *Health Aff (Millwood)* -2007;5:1399-406.
- 8 Sofaer S, Bryan S, Gold MR. Decision-makers' attitudes toward cost-effectiveness analysis. Abstract. Academy Health, Orlando, July 5, 2007.
- 9 Neumann PJ, Palmer JA, Daniels N, et al. Panel on integrating cost-effectiveness considerations into health policy decisions. *Am J Manag Care* 2008;14:185-8.
- 10 American College of Physicians. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med* 2008;149:956-61.