Health outcomes in economic evaluation: who should value health?

Eugena Stamuli*

Health Economics, York Trials Unit, Department of Health Sciences, University of York, York, UK

The valuation of health states is an integral part of economic evaluation studies. The source of these valuations (general public vs. patients) is surrounded by controversy. Health state values generated by the general public are often different compared with those of patients. General public values may not account for adaptation of the patients; patients’ values potentially incorporate self-interest. Decisions on the appropriate source of health values ultimately depend on the specific decision-making context and objectives of the evaluation. Differences in valuations and implications for decision-making should be explicitly addressed. Further research should systematically identify circumstances in which public and patients’ valuations diverge. There appears ground for development of methods that allow the simultaneous incorporation of patients and public preferences. Existing literature which discusses the issues around the health state valuation is summarized.

Keywords: health state valuation/preference/perspective/utilities/controversies/economic evaluation

Accepted: January 4, 2011

Introduction

The economic evaluation of healthcare interventions has become an integral part of health policy and decision-making for the allocation of the limited healthcare resources. For example, in Australia and Canada, economic evaluation is a mandatory requirement in applications for public reimbursement of new drugs. In the UK, the National Institute for Health and Clinical Excellence (NICE) incorporates assessments of cost-effectiveness when appraising medicines, medical devices and medical procedures. The basic principle of economic evaluation is that of assessing the ‘opportunity cost’, i.e. evaluating the benefits to patients that are foregone by using a resource in one way rather than its next best alternative. Those responsible for allocating resources need to prioritize between competing uses so that maximum benefit (health gain) can be obtained from any given
budget. Hence, choices have to be made about which services to provide in order to maximize patient benefit and minimize opportunity cost given a budget constraint; in other words, decisions about how to maximize the efficiency of the healthcare system. While the discussions around the appropriate perspective for these decisions and methods for the calculations of cost are still very vibrant in the field of health economics, this paper will focus only on the issues related to the measurement and valuation of health outcomes and the controversies surrounding these.

Before embarking on examining the controversies around the valuation of the health outcomes, or in fact what the valuation of health outcomes is, for use in economic evaluation, it is useful to give a description of the basic notions around health outcomes (as used in economic evaluation). A range of outcome measures are now used to evaluate healthcare technologies. These can be broadly divided into clinical and patient-based measures. Clinical measures are based on the physiological or pathological measurements made by clinicians and may have little relationship to patients’ feeling of well-being (e.g. lowered blood pressure for a hypertensive patient or a healed leg ulcer of a diabetic patient. These indicators are meaningful for the everyday clinical practice (ultimately, the clinician aims at healing the ulcer of the diabetic patient) and usually there is no controversy around these measures; there is small scope for subjectivity in their measurements and it is usually the clinician who makes the decision on a successful (or not) outcome. However, for the purpose of economic evaluation, they would only be useful if the decision question has a very narrow focus and relates to issues of productive efficiency\(^1\) e.g. which one of two interventions for the same health condition (bandage or compression for diabetic ulcers) provides better value for money. This type of economic evaluation is conventionally labelled as cost-effectiveness analysis.\(^2\)

Patient-based outcome measures are more closely related to patients’ perceptions of illness.\(^3\) In terms of complexity, they range from the measurement of patients’ symptoms to a detailed assessment of the impact of a condition and its treatment on patients’ functioning and general well-being. The latter is often described as a measure of health-related quality-of-life (HRQoL). In a broader decision-making context, where the allocation of the resources needs to be made across specialties and disease areas, it is critical that the outcome measure is generic, i.e. that it has meaning outside the specific clinical area within which it is used. This requires that the health outcomes used in the economic evaluation to be defined in such a way that simultaneously take into account the clinical effectiveness (survival rates, healing rates etc.), the morbidity of the patients, which is reflected in their HRQoL,
and the preferences of the patients or society about the benefits (or otherwise) of any intervention. The most popular example of this is the quality-adjusted life year (QALY). The premise of the QALY is that healthcare seeks to generate two general forms of outcome, increased life expectancy and improved HRQoL, and the QALY measure captures both elements. The use of the QALY facilitates comparisons across different sectors of healthcare and can assist the establishment of the allocative efficiency. In other words, decisions can be made for the maximization of the health of the society as a whole given the restricted budget allocations.

A QALY is constructed based on two components: the length of time (over which QALYs are calculated) and the associated HRQoL. The length of time is the clinical hard outcome (e.g. remaining life expectancy or life-years gained). The more difficult data for QALY estimation are the values (often referred to as utilities, preferences or weights) to quality adjust the length of time. One of the most controversial issues around QALYs is the source of these values: should the values be elicited from the general public or from patients? In other words, whose values should count? This question is rooted in two different aspects; one relates to the practical issues due to different degrees of knowledge and experience of the particular health state between patient and healthy respondent and the other relates to a broader question of whose preferences should contribute to the decisions on resource allocation, and ultimately how these are implemented in practice. The aim of this paper is to summarize the issues and controversies surrounding these questions and to identify areas for developing research.

How are the weighting values generated

The generation of utility values involves the use of either direct or indirect (generic, preference-based) elicitation methods. The direct valuation techniques of health states include standard gamble (SG), time trade-off (TTO) and the visual analogue scale (VAS). Although each method uses a different technique, they share the common intention of establishing the relative value of a particular health state in relation to the states of perfect health and death. If these utility values are to be used in economic evaluation, they should represent the strength of an individual’s preference for different health states under conditions of uncertainty based on expected utility theory. The SG and TTO are considered to be ‘preference-based’ techniques as they require respondents to make some sort of trade-off between different health states and either length of life or an immediate risk of death.
However, both methods have been criticized for being cognitively demanding for the respondents. In contrast, the use of VAS is the simplest of the direct elicitation methods, but does not involve any notion of uncertainty or any trade-off; hence it is not considered a real preference measurement instrument and is not typically recommended for use in economic evaluation studies.

In addition to the above methods, there is an ongoing research in the use of ordinal methods for the elicitation of health state values. These methods typically include discrete choice experiments and ranking. They are still in experimental stages and the different approaches are still being developed with the aim to produce values comparable with the more traditional valuation methods. It is of value to note that the data produced from ordinal methods need appropriate manipulation to be transformed into cardinal values. That is, the ordinal data need to be anchored against an absolute state (such as death) so that the values produced can be considered real health state values.

In recent years, the effort of having to elicit a set of weights for each and every QALY-based evaluation has been avoided by the development of indirect or standardized generic-classifications of states of health that, in principle, can be applied to all clinical areas. These classification approaches also provide a predetermined, preference-based weight for each state of health. The most commonly used are EuroQol (EQ-5D), the Short Form 6D (SF-6D) and the Health Utilities Index (HUI). The design of these instruments is based on some sort of generic description of health by the use of attributes (such as mobility or anxiety) and levels (such as ‘a lot’ or ‘a little’). In general, these descriptive systems have between five and nine dimensions with 3–6 levels each. Any given health state is the result of the combination of different levels across all the attributes. For example, the EQ-5D defines states of health in terms of five dimensions: mobility, self-care, usual activities, pain and depression/anxiety. Each dimension has three levels that can be used to classify a patient at a given point in time to one of 245 states of health. On the basis of interviews with the general public, utility weights have been estimated for each of these states of health. Therefore, if prognoses can be described in terms of the various EQ-5D states, QALY profiles can be defined for various interventions without having to directly elicit weights.

Both direct and indirect evaluation methods rely on the construction of health state descriptions or vignettes. For the indirect elicitation methods, the descriptive systems, have a standardized format and are constructed based on the fundamental principle that they should be sufficiently generic to be used in a wide (preferably all) range of patients; hence these are referred to as ‘off-the-shelf’ instruments. The vignettes, on the other hand, are bespoke descriptions of health states.
that have either a narrative form or are structured as bullet points. Ideally, the choice of the content should be such that satisfies the criteria of practicality, reliability, validity and responsiveness.³

The above criteria can be applied and controlled for in the standardized descriptive instruments, leading to statistical advantages. While bespoke health descriptions are often considered to better capture the patients experience¹⁴ as they include more natural and global description of the health states, they suffer from potentially directing the responded through subtle signals, labelling of the disease or framing effects.¹⁵ On the opposite side to this, one source of criticism for the ‘off-the-shelf’ instruments is around the potential lack of sensitivity for specific disease areas, which has led to the development of preference-based, condition specific measures.¹⁶,¹⁷

Irrespective of the descriptive approach employed, it is important to note that what is being valued is health states or profiles rather than possible change to these health states.¹⁸ Furthermore, in estimating the utility weights for these health states, a key decision needs to be made regarding from whom these values should be elicited. The valuation of the health outcomes can be conducted with different, distinct parts of the population: those who are experiencing the disease in real time (the patients), the healthcare professionals, the carers and the general members of the society who might have some or no experience of the disease at any point in their life.

The evidence that different sources of valuation may lead to different utility values is considered in the next section, followed by a discussion of the key considerations surrounding the issues of whose values should actually be reflected in economic evaluation studies.

### Evidence on and reasons for the differences

The comparison between the valuations derived from the general public and those from patients has been extensively explored in the literature. The majority of the literature suggests that there does appear to be a difference between the health state values generated from general public compared with those generated from patients themselves, with patients’ values for their own health state usually being higher than the values elicited for equivalent health states from the general public.¹⁹ However, another review of studies comparing patient values with values obtained from at least one other rater group suggests that several studies have not found evidence of a clear difference between patients and general public valuations.⁵ Furthermore, the same review concluded that, where these differences were observed in particular studies, the direction of this difference was not always
consistent. In the case of higher patient values, these tended to apply to more severe health states rather than mild states, for which patients give lower values than the general public.

The discrepancies in the valuation results from patients and the general public arise due to a plethora of reasons; see, for example, (19) or (20). These relate either to the way the health states are described for public evaluation; the way people perceive and interpret the information or, indeed, the different cognitive and psychological pathways and adjustments that the evaluator (patient or healthy public member) go through prior to or during the evaluation process. The most intuitive reason provided for the differences in values lies in the discrepancies between the health state descriptions and the health state as it is experienced by patients themselves. Insinga and Fryback introduced the notion of ‘lack of scope’ to describe this phenomenon. They found differences between the values generated from respondents’ valuation of their own, current health in the way that they perceived it and the equivalent EQ-5D health states used to describe their health. The values generated from their own health valuation were higher than those generated from the general public by using the EQ-5D scenarios. The conclusion was that EQ-5D scenarios may be too sparse to fully capture all the dimensions of certain health states. This potential criticism is not necessarily constrained to the use of ‘off the shelf’ classification systems. Even in the case of the descriptions based on vignettes or bespoke scenarios, which may be considered relatively complete descriptions, it might still be the case that they fail to comprehensively capture all aspect of HRQoL which will implicitly influence the valuation by healthy respondent. For example, although negative effects of being in a health state are numbered, the probabilities of these happening or the degree of impact on the individual are often not clearly stated.

The above, in conjunction with how the healthy responded will interpret or judge the given information, potentially impacts upon the elicited values. Ubel et al. argue that even when the descriptions of the health states under evaluation are identical, patients and the public are likely to value them idiosyncratically, with information based on their own experience or how they perceive things based on personal beliefs and stereotypes. When healthy people are asked to imagine being in a specific, impaired health state, they tend to focus more on negative aspects of the disease without taking into account the positives of someone’s life, which is a clear manifestation of ‘focusing illusion’ heuristic. Effectively, the valuation process is defined by the ‘shock’ that the healthy participant experiences while thinking about the impaired health state by focusing on the negatives and the inability to look at the bigger picture of life; it is their life with the disease that
patients value, rather than their health with the disease. Imagining being in a hypothetical health state is a demanding cognitive task that unavoidably leads to omitting other important facts of life such as adaptation to new health states and coping mechanisms that patients develop.

Developing coping mechanisms and adaptation to new situations is an uncontroversial fact of life. People tend adjust to new states, involving health, at two levels: physical and psychological. The former relates to the coping mechanisms that people develop such as skill enhancement or activity adjustments. New skills are acquired or existing skills become more focused so that people can achieve the same goals as they did prior to the onset of the disease. In addition, people tend to shift the focus to new, different activities in life which would provide them with the same satisfaction or sense of goal achievement. On a psychological level, the adjustment may take place due to lowering expectations or shifting expectations in life, whereby people draw satisfaction and happiness by looking at life from a different perspective. The process has also been labelled as ‘response shift’ describing the adaptation of internal standards of measurement to the new health status quo. People will find value to life aspects that they would not prior to the impaired health state.

In the evaluation process, healthy respondents tend not to recognize or predict adaptation to an impaired health state although one could argue that they could potentially consider what steps are needed to be made for adjustments in order to cope with the illness or disability. However, there appears to be no evidence that this happens in practice. On the opposite side of the argument, one question arises on whether the lack of acknowledgment of the adaptation from the healthy respondents is indeed important and whether it leads to fundamentally different results. Dolan argues that since the adaptation is a gradual and continuous process, the different time-points (early into the disease or later when the patients have developed coping mechanisms) could yield different results. Hence, it is not only a matter of making the healthy respondent aware of the adaptation of patients but also a matter of timing on when the values from the two sources are being compared. Consequently, eliciting values from a patient at the very early stages of the disease might be comparable with values obtained from a healthy respondent.

Irrespective of the timing of the evaluation, one has to bear in mind that by asking patient and healthy respondents, they receive two different perspectives. Kahnemann describes the former as a ‘buyer of health’ and the latter as a ‘seller of health’. This is a very important distinction to make while conducting valuation exercises since they involve some sort of trade-off between quality and length of life; ‘the
buyer’ and ‘the seller’ of health will attach very different values to their goods. This is in line with the now well understood phenomenon whereby the maximum amount one is willing to pay for a good (health, in this instance) differs significantly from the amount one is willing to accept to forgo that good.\textsuperscript{26} This is clearly demonstrated in the literature where people with severe illness are reluctant to trade-off life years.\textsuperscript{22} These perspectives need to be considered when interpreting the results from studies reporting different sources of valuations.

The reasons for differences between the valuations derived from the general public and from patients are summarized in Table 1.

### Table 1 Summary of reasons for the differences in health valuations

<table>
<thead>
<tr>
<th>Issues</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of health states</td>
<td>Health state description might not fully capture the patient’s experience of a specific health state, due to lack of scope</td>
</tr>
<tr>
<td>Perception and understanding of health state descriptions</td>
<td>People interpret health state description idiosyncratically. There is a tendency to focus on the negative aspects of the health state</td>
</tr>
<tr>
<td>Valuation of hypothetical health states is a cognitively demanding task that may lead to the omission of key aspects of a health state</td>
<td></td>
</tr>
<tr>
<td>Coping mechanisms and adaptation</td>
<td>People tend to adjust physically and psychologically to a particular health state which may not be recognized by a healthy respondent</td>
</tr>
</tbody>
</table>

The previous section has revealed that health state valuations provided by patients and by the general public often differ due to several different factors. Faced with these potential differences in values, a decision needs to be taken regarding from whom these should be elicited in the context of economic evaluation studies. To answer this question, aside from the practical issues described in Section 3, a key consideration is the decision-making context and the purpose of the analysis, since ultimately these should help to identify the appropriate source of the health state values. The decision-making context could be at a micro-level (such as the day-to-day clinical care or the choice of private health insurance) or at a broader, system-level where decisions need to be made concerning the allocation of health resources at a national level. In the first context, there appears little controversy on the view that it should be the patients’ own values that should be counted.\textsuperscript{27} For example, patient’s values could count in a context of shared...

\textsuperscript{26} E. Stamuli

\textsuperscript{27} British Medical Bulletin 2011;97
decision-making between the clinician and the patient.\textsuperscript{28} This view might be appropriate even when comparing interventions for the same condition;\textsuperscript{29} values of the patients experiencing that condition can be used for decision-making.

At a more macro- or system-level, when decisions concern the public good, the views on whose values should count are more divided. Two overarching reasons for this division are typically provided: the nature of the funding of the healthcare system and the political beliefs about the source of the welfare of the population. The choice of patients or their careers as source of preference values is in line with the ‘welfarist’ school of thought in economics which considers that the affected group of individuals should be the primary source of the health state values. The welfarist approach considers that individuals are the best judges of their own welfare, expressed by their individual utilities, the sum of which defines the overall welfare of the society. Hence, from a welfarist approach it should be the individuals affected by the condition who should provide the source of values. However, the assumptions underpinning this approach have been criticized in the context of the NHS, where individuals may not possess full information necessary to judge overall improvements in societal well-being. This is a pertinent issue to the NHS, where a fixed overall budget inevitably means that decision affecting one group of patients will have a knock-on effect on healthcare provision for another group of patients (who may or may not be identifiable). These issues have led to the development of an ‘extra-welfarist’ approach, which considers that any number of stakeholders, such as social decision makers (hence labelled as ‘decision maker’s approach’\textsuperscript{30}) or an average tax-payer, might be regarded as a more appropriate source of values than individual patients. The former is regarded as ‘individualistic’ approach, while the latter as ‘inherently paternalistic’.\textsuperscript{30}

Given the entirely publicly funded nature of the NHS, NICE currently adopts the ‘decision maker’s approach’. In its ‘Guide to the methods of technology appraisal’,\textsuperscript{31} the Institute advocates that:

For the reference case, the measurement of changes in the HRQoL should be reported directly from the patients and the value of changes in patients’ HRQoL (that is utilities) should be based on public preferences using a choice based method... The valuation of changes in HRQL reported by patients should be based on public preferences elicited using a choice-based method in a representative sample of the UK population.

Along these same lines, the panel on methods for economic evaluation in health advising the US Public Health Service recommended that fully informed representative sample of the general public should be the
appropriate source of health state values. This decision was based on the argument that the public bears the costs associated with healthcare decisions, hence they should be part of the decision-making process for the allocation of benefits.

Philosophical discussions advocating the ‘decision-maker’s approach’ provide additional arguments. Barry argues that ‘the reason for social preferences is to acquire some social state where the competing needs and claims of all members of the society are unified so as to reflect a common agreement between all the members of the society and satisfy the principle of justice.’ The principle of justice is satisfied additionally by employing the argument on the ‘veil of ignorance’ whereby the general public will value health states in an unbiased manner and blind to their own benefit, since they are not experiencing the impaired health state. Patients, on the other hand, could make subjective valuations with the ultimate aim to benefit from any new interventions. The ‘veil of ignorance’ coupled with questioning the general public in an ex ante context (i.e. there is uncertainty on whether or not the healthcare will be needed in the future) provides an even stronger basis for impartial decisions because not only will the respondents not have biased view about the benefits of treatments but they are also not certain whether they will even be in need of these interventions, reducing the potential for bias further.

The case against the public valuation of health states is built on the argument that they have limited or no experience of the health state under evaluation. The only true preference can be expressed by the patients as they have a real experience of the health state. Furthermore, the public valuations may yield biased results as what they value is not the true health state; what is valued is the fear and the shock of imagining how it is to live with the specific impaired condition due to ‘focusing illusion’ heuristics as discussed above. People tend not to look at the bigger picture of the patient’s life and the adaptation process they go through, which in turns, lead to overestimation of losses associated with the impaired health states.

On the opposite side of the argument, allowing patients to provide the values for their own health states can potentially compromise their potential to benefit from a new intervention. As discussed in Section 3, the evidence shows that patients tend to assign higher utility values to their own health state than the public, shaped by both lower expectations and adaptation. The argument against patient values is even stronger in the case when cognitive denial of patients suffering from a disease or suppressed recognition of full health are the factors that lead to adaptation or lower expectation. By assigning higher values to an impaired health state, the space for improvement (i.e. gaining QALYs) is small; hence any new treatment could be rejected based on the
non-effectiveness argument. The same argument goes the opposite direction, i.e. patients may overvalue the benefits of a new treatment with the ultimate aim to benefit from its implementation. On the contrary, skill enhancement, activity and goal adjustment or altered conceptions of health favour using the patients’ values since general public fail to take adaptation into account. Ultimately, the question is not whether adaptation is considered relevant in the decision-making process, rather the extent to which it is considered relevant per se.

Indeed, one argument builds on the fact that any new treatment will be provided to prospective patients; hence adaptation is not relevant. The question in this instance is not about whose values count, rather about how these values should be weighted given that the patients and/or member of public might have different degrees of experience and anticipation of illness.

The arguments for and against the use of patients or general public for the valuation of the health states have been summarized in Table 2.

### Table 2 Summary of arguments for and against the use of patients or general public to provide health state valuations

<table>
<thead>
<tr>
<th>Case</th>
<th>Valuation of health states by general public</th>
<th>Valuation of health states by patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>For</td>
<td>Individuals may not possess full information in the context of a decision-making for allocation of resources Satisfies the principle of justice Veil of ignorance (unbiased judgement due to lack of vested interest)</td>
<td>Patients are better judges of their own well-being (welfarist approach) They have full knowledge and experience of the particular health state</td>
</tr>
<tr>
<td>Against</td>
<td>Limited or no experience of the health state under evaluation Focusing illusion (overemphasis of negative aspects of health states) Fail to take adaptation into account</td>
<td>Patients may provide biased values because of vested interests or, They may compromise their own benefits by assigning higher values to health states due to adaptation</td>
</tr>
</tbody>
</table>

### Areas timely for developing research

While examining the current methods for obtaining health state values and the discrepancies between the values derived from the general public and the patients, it becomes evident that adopting either approach could lead to differing valuations. Currently in the literature, new avenues for developing methods that incorporate simultaneously the patients and general public values are being explored. For example, the case for obtaining better informed general population preferences by providing more information on patient’s experiences and adaptation has been advocated. This can be achieved by a variety of means such
as giving more information to the general public about what it is actually like to be living in the health states being valued through e.g. patient descriptions of their health condition in videos, patient-enhanced health state descriptors. Or, giving the general public information about patient values for identical health states and seeing if their values change as a consequence.

Conclusions

This review has served to identify the main issues and controversies surrounding the measurement and valuation of health outcomes in economic evaluation studies. The question of ‘whose values should count’ remains one of the most debated areas of health economics. Although no firm consensus exists regarding the most appropriate approach to both measure and value these outcomes, it has been argued that there is unlikely to be a wrong or right answer as different approaches and perspectives may lead to legitimate outcomes, depending on the decision-making context which is set by political views and strategies.

Identifying those healthcare interventions with the greatest potential for improving health remains a fundamental challenge for health policy decision makers. Inevitably, the current methods employed for the economic evaluation cannot be considered to provide a perfect solution to this problem. However, these approaches can help to inform the decision-making process by providing an explicit, transparent and consistent approach to decision-making. As the use of economic evaluation studies continues to grow, it is important that patients, clinicians and decision makers, as well as the general public, understand the key value judgements which lie behind these assessments.

Acknowledgement

Thanks to David Torgerson for the encouragement and support to write this paper.

References


