Health Economics, Policy and Law (2008), **3:** 69–77
© Cambridge University Press 2008 doi:10.1017/S1744133107004355

Debate

Developing methods that really do value the 'Q' in the QALY

PAUL DOLAN*
Tanaka Business School, Imperial College, London, UK

Abstract: Most health economists recommend that improvements in health be valued by asking members of the general public to imagine themselves in different states of health and then to think about how many years of life they would give up or what risk of death they would be willing to accept in order to be in full health. In this paper, I argue that preferences are not a very good guide to future experiences and a more suitable way to value health is to ask people in different states of health how they think and feel about their lives. Valuing health in this way may result in greater priority being given to mental health services. Whatever the precise implications, it is my contention that it is much better to ration health care according to real experiences rather than according to hypothetical preferences.

Introduction

Most of us would agree that decisions about who gets what treatment should be informed by the value of the benefits that health services generate. The question is how to judge the value of those benefits. Up until about 100 years ago, economists would have thought about benefits in terms of people's experiences – the greater an individual's enjoyment of an outcome, the greater the benefit. More recently, they have thought about benefits in terms of preferences – the stronger an individual's preference for that outcome, the greater the benefit. When it comes to valuing things – like health – that are not bought and sold in the market place, economists generally ask hypothetical questions about what an individual would be willing to pay in order to receive a particular benefit, such as a reduction in their risk of death or injury.

*Correspondence to: Paul Dolan, Professor of Economics, Tanaka Business School, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Email: paul.dolan@imperial.ac.uk

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

Methods have also been developed for valuing states of health that reflect respondents' willingness to exchange extra years of life or the risk of death for improvements in health. The unit of the scale that is used to evaluate this is a quality-adjusted life year (QALY), which assigns a weight between 0 (for death) and 1 (for full health) to each state of health and then multiplies that value by how long the state lasts. So, one QALY is equivalent to one year of life in full health. Some health care systems are now using QALYs to help determine priorities. For example, the National Institute for Health and Clinical Excellence (NICE), which offers guidance on the use of new and existing medicines and treatments within the NHS in England and Wales, has a 'strong preference' for expressing health benefits in terms of QALYs (www.nice.org.uk). There are, of course, many fairness-related factors besides health benefits that may be taken into account when determining priorities, but they are not the focus of the present paper, which concentrates on the valuation of benefits.

There are three questions that need to be addressed to calculate the 'quality adjustment' part of the QALY: *what* is to be valued; *how* is it to be valued; and *who* is to value it? The choice of 'what' refers to the aspects or dimensions of the health or well-being state being considered. There are two main ways health economists (and NICE) have addressed the question of 'how' we should value a particular state of health, and both methods rely on eliciting preferences. First, there is the time trade-off method (TTO). This requires respondents to consider how many years in full health is equivalent to a longer period of time in a poor health state. Second, there is the standard gamble method (SG). This requires respondents to consider the probability mix of full health and death that makes them indifferent between this gamble and the certainty of poor health. The choice of 'who' values a particular state essentially involves a choice between those with current experience of the health state in question ('patients') or those who are asked to imagine being in that state ('the public'). Many health economists (and NICE) have argued in favour of valuations from the public.

In this paper, I question the suitability of preference-based methods in calculating health state values. The focus of my critique is that the preferences of the general public are not reliable because the public are not good at assessing what it would be like to experience different states of health. The public tend to be biased in ways that lead them to overestimate the severity of a loss of health in certain conditions. This may suggest that preferences should instead be elicited from patients, as this would mean that the respondents would have direct experience of the health states in question. However, all responses to preference-based methods reflect whatever the respondent focuses on at the time of the assessment, which may not be what they will focus on while experiencing that health state. Patients could be asked to consider their previous experiences when making choices about the future but we are not very much better at remembering past experiences than we are at predicting future ones.

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

To more accurately reflect the effect of different health states on people's well-being, I propose that policy-makers in health and elsewhere should shift their attention from the measurement of preferences towards the measurement of experiences. There are problems with this approach too, particularly in relation to generating QALYs, but they are not insurmountable and I suggest that future research efforts should be directed towards generating experience-based QALYs for use by NICE and others in making resource allocation decisions.

Problems with preferences

Consider the following health state: 'some problems walking about, no problems washing or dressing self, no problems performing usual activities, no pain/discomfort, not anxious/depressed'. This is one of the health states defined by the EQ-5D classification system which NICE currently considers to be 'the most appropriate choice in the UK'. Imagine being asked to choose between living for ten years in this state and then dying, or for a shorter period of time in full health. Over 3,000 members of the general public were asked precisely these kinds of preference-based questions (Dolan, 1997), the data from which have been used in a number of submissions to NICE. The results show that the average person considers ten years with some problems walking about as equal to 8.5 years in full health. As such, the quality adjustment weight for QALYs is taken to be 0.85, or a 15% loss from full health.

Now, almost all accounts of well-being that are based on preferences require that people make full use of all relevant information (Harsanyi, 1996). In this context, I suggest that 'relevant information' should include an understanding of the likely future experiences associated with the different choices. If you are likely to adapt to your walking problems over those ten years – and there is good evidence that you would adapt – then your valuation should reflect the use of this relevant information. The evidence currently available suggests that members of the public tend to overestimate the losses associated with a range of health states (de Wit *et al.*, 2000).

There are at least three factors that tend to inflate the public's assessments of the severity of hypothetical health states (Dolan and Kahneman, 2008). First, due to the nature of preference-based questions, the respondent's attention is drawn to the transition from one state to another. Initially, paraplegia, for example, will have a big effect on well-being but, once a person has been paraplegic for a period of time, they tend to withdraw their attention from their paraplegia and focus on the many other things in life, and so their well-being improves over time. Therefore, valuations are likely to be affected by a 'Peak–Start Rule' where respondents focus on the worst effects of a health change and the effects that are experienced immediately (Dolan and White, 2006). For many adverse conditions, the peak and the start will coincide.

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

Second, the respondent's attention is focused on the health domain rather than on other domains (such as personal relationships) which may be unaffected, or even enhanced, by changed health status. Health states are typically described using only a limited number of dimensions and always in ways that draw the respondent's attention to those dimensions that will be adversely affected. However, one of the ways in which we deal with changed circumstances is by redeploying our attention. So, if adaptation to paraplegia takes the form of not thinking about it (but rather thinking about domains of life other than health), then focusing respondents' attention on the health domain will lead them to overestimate the impact of paraplegia.

Third, it is possible that responses to preference-based questions reflect immediate emotional reactions to the health state in question, which in the case of some severe health states is likely to be an initial shock reaction to, or fear associated with, that state. So, not only might respondents be channelled to consider a limited number of (possibly relatively unimportant) aspects of the future, they might even be channelled away from thinking about the future at all, and towards focusing on current feelings. Of course, policy-makers may wish to devote resources to the health states that people fear the most, but accounting for fear is a quite separate issue from accounting for the losses in well-being from a given health state, and preference-based valuations conflate fears that people have about experiencing poor health with their assessments of how their lives will be affected by poor health.

One way to try and solve the problems associated with eliciting values of hypothetical health states from the public is to elicit the preferences of those currently experiencing the health state in question. However, the fundamental problem with these valuations remains – the responses cannot avoid focusing effects. Patients may well have experience of the state they are being asked to value but they cannot possibly be expected to weigh up all their possible future experiences when considering their responses. In addition, methods such as the TTO and SG require patients to consider how their future experiences would be different were they to be in full health. Whilst many patients would have had previous experience of full health, their recollection of this – at least in an evaluative sense – may be far from perfect and they are again likely to focus on only a limited number of ways in which their lives would be different from now.

Even if patients were able to accurately forecast how their current health would affect them in the future, their preferences are unlikely to reflect how that state had affected them in the past. Patients are likely to use their current preferences to rationalize a previous change, so they may well remember the transition into their current health state as being less intense than it was felt to be at the time. There is now good evidence that the retrospective recall of health is highly correlated with current health state and not so well correlated with the initial state. More generally, our memories do not recall past experiences and their duration particularly well: rather, we use a 'Peak–End Rule',

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

which focuses on the most intense and most recent experience and ignores the full set of experiences and their associated durations (Kahneman *et al.*, 1997).

Valuing experiences

It may be possible to develop more sophisticated preference-based methods and to provide respondents with more information about the experiences associated with the states they are asked to value. Ubel *et al.* (2001) devised a number of novel studies in which they attempted to draw respondents' attention away from the negative effects of different hypothetical conditions by asking them to consider how each condition would affect a range of different domains of life. However, in only one out of ten questionnaire variants did respondents' valuations of the health states increase, suggesting that it is difficult to get respondents to focus on things other than negative consequences.

Therefore, I suggest that we look for more direct measures of the experiences associated with different states of the world. Respondents would be asked to rate their subjective well-being (SWB) without drawing their attention to any aspects of their life, such as their health. We can then ask them to describe their health in some way, perhaps even using a generic classification system as recommended by NICE, and to provide information on a range of other factors that are known to be associated with subjective well-being (income, marital status etc.). By controlling for these other factors, we will then be able to estimate the effect that different health states have on subjective well-being. Allowing statistical analyses to determine the relative weights attached to the various factors that affect SWB seems a far more sensible approach than relying on what an individual thinks these weights should be.

SWB is usually measured by asking individuals to state how satisfied they are with their lives overall, usually on a scale between 1 and 10. Economists and policy-makers have been showing increasing interest in the use of measures of subjective well-being to inform public policy (Dolan *et al.*, 2008). In part, this interest stems from the availability of large longitudinal datasets which enable us to show how an individual's own happiness changes in response to changes in health etc. over time. One such survey is the British Household Panel Survey (BHPS), which is an in-depth longitudinal survey of a nationally representative sample of about 5,000 households (10,000 individuals) in the UK. The survey has been running since 1991 and, since 1996, all respondents have been asked how satisfied they are with their life overall (on a 1–7 scale).

In the 1999 and 2004 BHPS, respondents were also asked to describe their health using a generic classification system (the SF-36), and so we are able to show how the change in an individual's health across two waves of the survey is related to how SWB changes across the same period. SG data for the SF-36 are also available from 600 members of the UK general public (Brazier *et al.*, 1999),

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

thus allowing for a comparison between preference-based and experience-based valuations. Details can be found in Dolan (2007) but the important point from a resource allocation perspective is that the priority accorded to different conditions would be different based on the two sets of weights. SWB ratings suggest that curing mental health problems is much more important than curing pain, whilst SG preferences suggest that pain is more important.

Problems with subjective well-being ratings

The reduction of focusing effects from the assessment of subjective well-being is one of its great advantages but it is difficult to think how they can be completely removed in those studies where respondents would be told that they are being asked such questions precisely because they have a particular health condition. In addition, responses to global life satisfaction questions will be based to some extent on whatever the respondent's attention is drawn to at the time of the assessment. Responses have been shown to be affected by more immediate concerns, such as current mood, whether someone else is in the room or not, and even by the weather. On the other hand, there is evidence to suggest that global measures are relatively stable constructs, and correlate well with other indicators of how an individual's life is going, such as informant reports and even with reduced suicide attempts.

There are new developments in the measurement of well-being moment-by-moment, which provide an alternative to SWB ratings. The day reconstruction method (DRM) asks respondents to divide the previous day into a number of episodes and then to rate different feelings during those activities on a 1–6 scale (Kahneman *et al*., 2004). The DRM currently taps into how a person feels and not into how she thinks her life is going. In comparison, SWB ratings have the advantage that they can reflect whatever matters to the individual, e.g. the purpose and meaning they have in life, as well as how they feel. However, there is nothing in principle that prevents day reconstruction method type measures from measuring a richer set of experiences than feelings.

To produce data that allows the relative cost effectiveness of different interventions to be calculated requires SWB and DRM ratings to be expressed on a cardinal scale, where the change from 1 to 2 on the scale would be the same as the change from 3 to 4. SWB responses are normally analysed as ordinal data but treating responses as cardinal gives very similar results. There are different ways in which the ratings of different feelings in the DRM can be aggregated and the sensitivity of various aggregation rules should be tested. Importantly, though, any summary measure of feelings will have cardinal properties so long as each unit of time is treated equally.

QALYs currently express health on a scale between death and full health. For many policy applications, and since almost everyone would not rate their current

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

state as worse than dead by any measure of experience, we can estimate the loss in well-being from premature death to be the well-being that would have been expected otherwise. However, there will be other applications, such as the evaluation of short-term acute conditions, where a valuation for death will be necessary, and so future studies should calibrate SWB ratings against an explicit lower anchor of death (I really do not see any need to estimate values for states worse than dead). The top anchor for SWB ratings does not have to be full health – in fact, it would be better if it were not framed in terms of health at all – but it does have to be something that allows for comparison across studies.

One potentially important problem that would limit the feasibility of comparing results across studies is response shift (Sprangers and Schwartz, 1999). For example, those reporting pain in the BHPS might compare their SWB to others in pain, which would lead to over interpretation of the degree of any adaptation. Whilst response shift is certainly an important area for future research, it certainly cannot explain all the changes in preferences and experiences that take place. For example, there is evidence of adaptation when physiological measures are used, which should be less prone to response shift, e.g. Dar *et al.*, 1995 found that war veterans with more severe past injuries could hold their finger in hot water for longer before classifying it as painful than veterans with less severe past injuries.

There are more fundamental concerns about the nature of adaptation itself. All else equal, the more a person adapts to her condition, the less priority she will receive in the competition for resources that improve quality of life. In light of any costs associated with adaptation, it may be unjust if some patients lost out in the race for resources because their effort diminished the value of treatments for them compared to other patients who did not expend the same effort. However, if this argument is accepted, then an advantage gets created for those who have adapted. Treatments for such people will get greater priority than are warranted by the size of the actual benefit from them. Resource allocation decisions will then be made as if an adapted person's gains in well-being count for more than another less well-adapted person's gains. This also seems unfair. So, for the most part, it seems legitimate to give greater priority to those conditions for which the adaptation process is long and/or incomplete – precisely because the adaptation process is long and/or incomplete.

Concluding remarks

We can think of public interventions as benefiting people if their preferences are satisfied or if their experiences are enhanced. The methods favoured by economists and by NICE reflect the former interpretation. To the extent that our wants, as captured by our decisions, are based on predictions of what we will subsequently enjoy, we are often guilty of 'miswanting', i.e. we want things that do

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

not make us happier and vice versa (Gilbert and Wilson, 2000). I was closely involved in the study that elicited TTO valuations for health states from the general public and the results from this study are looked on favourably by NICE.

Unfortunately, I cannot think of a better example of preferences that are guilty of 'miswanting'. The average valuation suggests that 'some pain or discomfort' is worse than being 'moderately anxious or depressed' when the limited evidence we have on SWB suggests precisely the opposite. The use of values like those I was involved in generating may therefore result in a gross misallocation of resources. Eliciting preferences from patients will avoid some of the problems associated with eliciting preferences from the general public but a patient's preference-based values may still not reflect the future well-being associated with their health state.

It is my contention that an individual's assessment of her life offers a more defensible measure of well-being than the satisfaction of her actual preferences, and measures of SWB may more closely reflect those preferences an individual would have if only she was faced with full information about the consequences of her choices. At a practical level, SWB questions are less cognitively demanding than preference-based questions, and they also have the major advantage that respondents do not need to be made aware of the particular health state being valued, therefore removing the risk of artificially created preferences.

Whilst the onus of responsibility for good quality data is removed from the respondents, it is placed on the analytical techniques used to determine the effect of particular health states on subjective well-being. There are data from existing surveys, such as the BHPS, which could be explored further to illustrate the effects of different health conditions on SWB. However, we also require large longitudinal studies that look at SWB pre- and post-intervention to determine the benefits from intervention. I hope that I have encouraged others to join me in a research endeavour that seeks to ration health care in ways that improves the real experiences of how people think and feel about their lives rather than in ways that satisfies their hypothetical preferences over how they imagine thinking and feeling.

Acknowledgements

This paper draws heavily on ideas developed with Daniel Kahneman when I was a Visiting Research Scholar at Princeton University. I am also grateful to Barney Gough, Rob Metcalfe, Tessa Peasgood, Ann Rossiter, and Mat White for various discussions. More detailed discussions of the issues raised here can be found in: Dolan, P. and Kahneman, D. (2008) Interpretations of utility and their implications for the valuation of health, *Economic Journal*; and Dolan, P. (2007) Finding a NICEr way to value health: from hypothetical

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press

preferences to real experiences, *Social Market Foundation*. Naturally, I am fully responsible for the views expressed here and they may well not reflect the views of any of the aforementioned colleagues.

References

Brazier, J., J. Roberts, and M. Deverill (1999), 'The estimation of a preference-based measure of health from the SF-36', *Journal of Health Economics*, **21**(2): 271–292.

Broome, J. (1993), 'Qalys', *Journal of Public Economics*, **50**: 149–167.

Dar, R., D. Ariely, and H. Frenk (1995), 'The effect of past injury on pain threshold and tolerance', *Pain*, **60**: 189–193.

Dolan, P. (1997), 'Modelling valuations for EuroQol health states', *Medical Care*, **11**: 1095–1108.

Dolan, P. (2000), 'The measurement of health-related quality of life for use in resource allocation decisions in health care', in A. J. Culver and J. P. Newhouse (eds), *Handbook of Health Economics*, Vol. I, Amsterdam: Elsevier Science BV, pp. 1724–1760.

Dolan, P. (2007), 'Finding a NICEr way to value health: from hypothetical preferences to real experiences', *Social Market Foundation*.

Dolan, P. and D. Kahneman (2008), 'Interpretations of utility and their implications for the valuation of health', *Economic Journal*.

Dolan, P., T. Peasgood, and M. White (2008), 'Do we really know what makes us happy: a review of the economic literature on the factors associated with subjective well-being', *Journal of Economic Psychology*.

Dolan, P. and M. White (2006), 'Dynamic well-being: connecting what we anticipate with what we experience', *Social Indicators Research*, **75**(2): 303–333.

Dolan, P., R. Shaw, A. Tsuchiya, and A. Williams (2005), 'QALY maximisation and people's preferences: a methodological review of the literature', *Health Economics*, **14**(2): 197–208.

Gilbert, D. T. and T. D. Wilson (2000), 'Miswanting: some problems in the forecasting of future affective states', in J. Forgas (ed.), *Thinking and Feeling: The Role of Affect in Social Cognition*, Cambridge: Cambridge University Press.

Harsanyi, J. C. (1996), 'Utilities, preferences, and substantive goods', *Social Choice and Welfare*, **14**: 129–145.

Kahneman, D., P. Wakker, and R. Sarin (1997), 'Back to Bentham? Explorations of experienced utility', *Quarterly Journal of Economics*, **112**: 375–405.

Kahneman, D., A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone (2004), 'Toward national well-being accounts', *American Economic Review*, **94**(2): 429–434.

Sprangers, M. A. G. and C. E. Schwartz (1999), 'Integrating response shift into health-related quality of life research: a theoretical model', *Social Science and Medicine*, **48**(11): 1507–1515.

Ubel, P. A., G. Loewenstein, J. Hershey, J. Baron, T. Mohr, D. A. Asch, and C. Jepson (2001), 'Do nonpatients underestimate the quality of life associated with chronic health conditions because of a focusing illusion?', *Medical Decision Making*, **21**: 190–199.

de Wit, G. A., J. Busschbach, and F. De Charro (2000), 'Sensitivity and perspective in the valuation of health status: whose values count?', *Health Economics*, **9**(2): 109–126.

https://doi.org/10.1017/S1744133107004355 Published online by Cambridge University Press