

Finding a NICEr way to value health: From hypothetical preferences to real experiences

Paul Dolan

INTRODUCTION

Spending on the National Health Service (NHS) in the UK increased as a percentage of GDP from 7.0% in 2000 to 9.4% in 2006.¹ However, even with this substantial increase in funding, the NHS cannot afford to provide every effective drug or treatment to every patient. Decisions still have to be made about what services should and should not be made routinely available on the NHS. Priority setting – or rationing (which amounts to the same thing) – is unavoidable. 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.

First published by
The Social Market Foundation,
July 2007

The Social Market Foundation
11 Tufton Street
London SW1P 3QB

Copyright © The Social Market
Foundation, 2007

The moral right of the authors has been asserted. All rights reserved. Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of both the copyright owner and the publisher of this book.

Design www.jadedesign.co.uk

The Social Market Foundation

The Foundation's main activity is to commission and publish original papers by independent academic and other experts on key topics in the economic and social fields, with a view to stimulating public discussion on the performance of markets and the social framework within which they operate.

The Foundation is a registered charity and a company limited by guarantee. It is independent of any political party or group and is financed by the sale of publications and by voluntary donations from individuals, organisations and companies.

The views expressed in publications are those of the authors and do not represent a corporate opinion of the Foundation.

Chairman
David Lipsey (Lord Lipsey of Tooting Bec)

Members of the Board
Viscount Chandos
Gavyn Davies
David Edmonds
Brian Pomeroy
Martin Ivens
Shriti Vadera

Director
Ann Rossiter

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, or pleasure from, an outcome, the greater her 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. So, 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.²

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. Health economists are fond

¹ Office of Health Economics (2007), <http://www.ohe.org/page/index.cfm>.

² For example see: *Green Book, Appraisal and Evaluation in Central Government* (London: HM Treasury, 2003), <http://greenbook.treasury.gov.uk/>.

of asking the general public questions like ‘How many years in full health would you consider equivalent to ten years in moderate pain?’ 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 healthcare 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. NICE recommends funding technologies with a cost-per-QALY gained less than £30,000.

When valuing a state of health 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’ is to be valued refers to the aspects or dimensions of the health state being considered.³ Condition-specific measures focus solely on the medical condition being valued, for example osteoarthritis or diabetes, and provide a detailed account of only those dimensions of health directly affected by the condition. Alternatively, generic measures cover a broader range of dimensions but at the expense of sensitivity on any one dimension. Condition-specific measures are suited to comparisons within conditions whilst generic measures are suited to comparisons across conditions.⁴

There are several ways that we can look at ‘how’ we should value a

³ Health dimensions refer to attributes of health such as physical, social and cognitive function, mood and pain.

⁴ Despite being referred to collectively as measures of ‘health-related quality of life’, the various generic health state classification systems differ widely in what is considered ‘health-related’ and what is considered ‘beyond

health’. Such issues, and those more generally concerned with the distinction between ‘health’ and ‘well-being’, are not considered further here.

particular state of health. I look at two methods currently approved by NICE here. First, there is the time trade-off method. This requires respondents to consider how many years in full health are equivalent to a longer period of time in a poor health state. Second, there is the standard gamble method. 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. These methods are preference-based because they require respondents to make choices between different future health prospects.

The choice of ‘who’ values a particular state of health refers to whom the valuation should be elicited from. Essentially this 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’).

In relation to these questions, NICE requires that health states ‘should be measured in patients using a generic and validated classification system for which reliable UK population preference values, elicited using a choice-based method such as the time trade-off or standard gamble (but not rating scale), are available’.⁵ So, whilst additional analyses may be submitted, NICE’s ‘reference case’ is clear: ‘what’ is a generic health state descriptive system, ‘how’ is a preference-based method, and ‘who’ is the general public.⁶

In this paper, I will highlight serious problems with NICE’s responses to the second and third questions – that is, ‘how’ a state of health should be valued and by ‘whom’.⁷ In the next section I will question the

⁵ National Institute for Clinical Excellence, *Guide to the Methods of Technology Appraisal* (NICE, 2004).

⁶ Many other countries (Australia, Canada, Hungary Italy, Netherlands etc.) have been developing guidelines for health technology assessments which broadly follow these recommendations.

⁷ A similar account of many of the arguments I make here can be found in Dolan, P. and Kahneman, D., (forthcoming). Interpretations of utility and their implications for the valuation of health, *Economic Journal*.

suitability of preference-based methods for the task of calculating the value of different states of health.

The focus of my critique is that the preferences of the general public are not a reliable basis for judgement because the general 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, rather than the general public, as this would mean that the respondents would have direct experience of the health states in question. However, all responses to preference-based methods, whoever the respondents are, reflect whatever the respondent focuses on at the time of the assessment, which may not be what they will focus on or attend to while experiencing that health state. Patients could be asked to consider their previous experiences when making hypothetical choices about the future but there is evidence that people are not very much better at remembering past experiences than they are at predicting future ones.

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

existing data from research into subjective well-being – how individuals think and feel about their lives – that could be used in this regard and, in section three, I will present some fresh evidence to show that using subjective well-being based quality adjustment weights for QALYs may have different implications for resource allocation decisions than using preference-based weights. In section four, I will present some of the criticisms against using subjective well-being, and will argue that many of these criticisms have been overstated. In section five, I provide some concluding remarks.

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’.⁸ 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. I was involved in a study that asked over 3000 members of the general public precisely these kinds of preference-based questions,⁹ the data from which have been used in a number of submissions to NICE. The results show that the average person considers 10 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 the respondents making those preferences make full use of all relevant information.¹⁰ 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 probably would, then your preference-based valuation should presumably reflect the use of this relevant information. The evidence currently available suggests that members of the general public are not aware of all of the relevant information, as they tend to overestimate the losses associated with a range of health states. In a review of 39 studies, de Wit et al¹¹ found that 23 studies report patient values to be higher than public ones, 2 report

public values to be higher, 11 report no difference, and 3 report contradictory findings. In other words, patients rate their own health state more positively than the general public does. Some of the differences could have important implications for resource allocation decisions.¹²

There are at least three factors that tend to inflate the public’s assessments of the severity of hypothetical health states. First, due to the nature of preference-based questions, the respondent’s attention is drawn to the transition from one health state to another. If we take paraplegia, for instance, getting used to being paraplegic is extremely difficult, and it is this transition that is focused on in hypothetical questions. However, 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. So, valuations are likely to be affected by a ‘Peak-Start Rule’,¹³ where respondents focus on the worst effects of a health change and the effects that are experienced immediately. The worst effects of most health changes will be experienced immediately, and so the peak and start are likely to coincide and to lead respondents to almost ignore all longer term (usually less intense) consequences.

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

8 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’. The EQ-5D defines health in terms of five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension contains three levels, broadly defined as ‘none’, ‘some’ and ‘extreme’ problems.

9 Dolan, P. (1997). Modelling valuations for

EuroQol health states. *Medical Care*, 11, 1095-1108.

10 See for example: Harsanyi, J.C. (1996). Utilities, preferences, and substantive goods. *Social Choice and Welfare*, 14, 129-145.

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

12 For example, in the first empirical study on this issue, Sackett and Torrance (1978) asked the public and home dialysis patients to value a remaining lifetime with chronic dialysis using the time trade-off (TTO), preference-based, method. The average utility from the general public was 0.39, as compared to 0.56 from dialysis patients. If it were possible to return dialysis patients to full health, then the gain

would be 0.61 QALYs per year (1.0-0.39) if we used public utilities and 0.44 (1.0-0.56) if we used patient utilities. In this case, using public values would reduce the cost-per-QALY of organ transplantation by nearly 40%.

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

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. Insofar as fears surrounding particular health states can be reduced by expenditure on the treatment and prevention of those states, policy-makers may be able to increase well-being by spending in such a way. Moreover, the public, through the democratic process or pressure groups, may demand that certain health states be given greater priority.

However, 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. I am not aware of anyone arguing for the preference-based models on the grounds that they pick up people's legitimate emotional responses to the health states in question. Indeed, economists have described health in terms of dimensions of health, such as mobility, rather than in terms of conditions, like cancer, in order to avoid introducing too much emotion into the responses. Preference-based models are assumed to tap into people's cognitive

assessments of the utility associated with those states, despite the lack of any evidence to support such an assumption.

One way to try and solve the problems associated with eliciting values of hypothetical health states from the public is to calculate the value of a state of health by asking patients who are currently experiencing that health state about their preferences. However, the fundamental problem with these valuations remains – namely, that the responses cannot avoid focusing effects. Even if the respondent is a patient who has experience of the health state in question, their responses will reflect whatever their attention is drawn to at the time of the assessment, rather than what it will be drawn to in future experiences. 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, both the preference-based methods introduced above 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. Different patients may focus on different aspects of their future lives and may have very different conceptions of what full health would mean to them. These differences could go some way towards explaining the wide variation in preferences elicited from reasonably homogenous groups of patients.

Even if patients were able to accurately forecast how their current health would affect them in the future, their preferences would be unlikely to reflect how that state had affected them in the past. Just as the general public will use their current preferences to rationalise a future change, patients are likely to use their current preferences to rationalise 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 an individual's current health state and not so well correlated with her initial state. More generally, our memories do not recall past experiences and their duration particularly well. Rather, we use a 'Peak-End Rule', which focuses on the most intense and most recent experience and ignores the full set of experiences and their associated durations.¹⁴

¹⁴ Kahneman, D., Wakker, P. and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, 112, 375-405.

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. Many studies now gather qualitative data that seeks to 'get behind the numbers' in ways that help us to understand the thought processes respondents use to generate responses. Armed with such data, we can design better questions to generate the answers we need for policy appraisal. However, preferences will still reflect whatever a respondent's attention is drawn to at the time of the assessment. Ubel et al 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.¹⁵ However, in only one out of ten questionnaire variants did respondents' valuations of the health states increase as a result, 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. This could mean that valuations would be elicited using something like a rating scale,¹⁶ which Broome advocated should be used to rate the 'goodness' of a health state.¹⁷ Of course, such a rating scale has been explicitly ruled out by NICE and is not favoured by economists because it does not require respondents to make a trade-off between

¹⁵ Ubel, P.A., Loewenstein, G., Hershey, J., Baron, J., Mohr, T., Asch, D.A. and Jepson, C. (2001). Do nonpatients underestimate the quality of life associated with chronic health conditions because of a focusing illusion? *Medical Decision Making*, 21, 190-199.

¹⁶ A health state rating scale consists of a line with anchors representing perfect health and

death. Respondents place each health state being rated on the line at the point where they believe the health state lies, between death and perfect health. The intervals between the placements reflect the perceived differences between the health states.

¹⁷ Broome, J. (1993). *Qalys*. *Journal of Public Economics*, 50, 149-167.

health and something else of value to the individual.¹⁸ However, the rating scale I have in mind would ask respondents to rate their overall subjective well-being without drawing their attention to any specific 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 well-being seems a far more sensible approach than relying on what an individual thinks these weights should be. As mentioned above, an individual's judgement on these matters is affected by focusing effects and a range of errors, both in forecasting well-being in a hypothetical future health state and recalling past experience of their own health state. Of course, to calculate the effect of health states on subjective well-being requires us first to find a reliable way of measuring that well-being. In most of the studies that I am aware of, subjective well-being is usually measured by asking individuals to state how satisfied they are with their lives overall.¹⁹

Economists and policy makers have been showing increasing interest in the use of measures of subjective well-being to inform

18 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 Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1, pp. 1724-1760). Amsterdam: Elsevier Science B.V.

19 For example, the British Household Panel Survey asks 'Here are some questions about

how you feel about your life. Please tick the number which you feel best describes how dissatisfied or satisfied you are with the following aspects of your current situation... your life overall?' The response scale is 1 to 7, anchored at not satisfied (1) and completely satisfied (7).

public policy.²⁰ In part, this interest stems from the availability of large longitudinal datasets that 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, 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.

In the 1999 and 2004 British Household Panel Surveys, respondents were also asked to describe their health. The surveys showed that the change in an individual's health across two waves of the survey was related to how his subjective well-being changed across the same period. The results of our analyses suggest that the biggest effects on this well-being are associated with changes in mental health and

20 For a recent review of all papers in economics journals since 1990 on the determinants of happiness, see: Dolan, P. and White, M. (2006). Dynamic well-being: connecting what we anticipate with what we experience. *Social Indicators Research*, 75, 2, 303-333.

vitality, and that the smallest effects are associated with pain and physical role limitations. These results should only be treated as illustrative of the relative importance of the dimensions, as much depends on the way the different levels within each dimension are described.

A much more interesting comparison in the context of this paper is between the value attached to different dimensions of health when calculated from analysis of subjective well-being ratings and the value attached to them from the preferences of the general public. Using the (preference-based) standard gamble method, values for health states were elicited from 611 members of the general public.²¹ The quality-adjustment weights from this study would be looked on very favourably by NICE as they satisfy all the requirements of their ‘reference case’ (generic classification system, preference-based valuation method and population values). Table 1 shows the QALY loss associated with the worst level of each of the health dimensions, based on subjective well-being ratings from the British Household Panel Survey²² and standard gamble responses from the public.

The values of the losses in quality of life on five of the dimensions are greater when based on standard gamble responses as compared to subjective well-being ratings. The large differences on two of these dimensions (social functioning and pain) are consistent with the general public overestimating severity, although for mental health the value of the quality of life loss is slightly less according to standard gamble responses. 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. Subjective well-being ratings suggest that curing mental health problems is much more important than curing pain, whilst standard gamble preferences suggest that pain is more important. Whilst the weights attached to vitality are not much different, it is the second most important dimension according to subjective well-being, yet the second least important according to the standard gamble.

Table 1: Comparison of QALY losses based on subjective well-being and standard gamble preferences²³

Dimension and description of the worse level	QALY loss on a 0-1 scale based on subjective well-being	QALY loss based on standard gamble preferences	Difference (subjective well-being - standard gamble)
<u>Physical functioning</u> 'Your health limits you a lot in bathing and dressing yourself'	-0.071	-0.111	-0.040
<u>Role limitations</u> 'You are limited in the kind of work or other activities as a result of your physical health'	-0.039	-0.050	-0.011
<u>Social functioning</u> 'Your health limits your social activities all of the time'	-0.002	-0.087	-0.085
<u>Pain</u> 'You have pain that interferes with your normal work extremely'	-0.034	-0.167	-0.133
<u>Mental health</u> 'You feel tense or downhearted or low all of the time'	-0.159	-0.128	0.031
<u>Vitality</u> 'You have a lot of energy none of the time'	-0.089	-0.091	-0.002

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

²² This analysis controlled for income, age and marriage status, and, using fixed effects, for individual differences in the use of the life satisfaction scale.

²³ Taken from Brazier, J., Roberts, J. and Deverill, M. (1999). The estimation of a preference-based measure of health from the SF-36, *Journal of Health Economics*, 21, 2, 271-292.

PROBLEMS WITH SUBJECTIVE WELL-BEING RATINGS

The reduction of focusing effects in 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 life-as-a-whole or global life satisfaction questions²⁴ 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 reasonably well with other indicators of how an individual's life is going, such as informant reports and even reduced suicide attempts.

There are new developments in the measurement of well-being moment-to-moment which provide an alternative to subjective well-being ratings. The day reconstruction method asks respondents to divide the previous day into a number of episodes and then to rate different feelings during those episodes on a 1-6 scale.²⁵ It is possible to look at the effect of different health states on a moment-to-moment basis, as well as to consider whether different states also have an effect on the activities that people engage in. The day reconstruction

method currently taps into how a person feels and not into how she thinks her life is going. In comparison, subjective well-being ratings have the advantage that they can reflect whatever matters to the individual, for example, the purpose and meaning they have in life, as well as how they feel. However, there is nothing in principle that prevents day reconstruction type measures from measuring a richer set of experiences than feelings.

To produce data that allow the relative cost-effectiveness of different interventions to be calculated (e.g. as expressed by cost-per-QALY ratios) requires subjective well-being ratings and the day reconstruction method to be expressed on a cardinal scale, where the change from 1 to 2 on the scale is the same as the change from 3 to 4. Subjective well-being responses are normally analysed as ordinal data (that is, 2 is better than 1 and 4 is better than 3 but we do not know by how much) but treating responses as cardinal gives very similar results. There are different ways in which the ratings of different feelings in the day reconstruction method can be aggregated. Kahneman and Krueger, for example, propose the 'U-Index', which looks at the feeling that gets the highest rating (if this is a negative feeling, the time in that activity is scored as one; otherwise it is scored as zero) and then calculates the proportion of time that people spend in an unpleasant state.²⁶ The sensitivity of various aggregation rules should be tested, but any summary measure of feelings will have cardinal properties so long as each unit of time is treated equally.

²⁴ Global life satisfaction (unlike life-domain satisfaction, which refers to satisfaction with specific areas of an individual's life, such as work or income) refers to satisfaction with a broad conception of an individual's life.

²⁵ Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N. and Stone, A. A. (2004). Toward national well-being accounts. *American Economic Review*, 94, 2, 429-434.

²⁶ Ibid.

QALYs currently express health on a scale between death and full health. It is my sense that we do not need an endpoint of death. Since almost everyone would not rate their current 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. I appreciate that this does not allow for states rated as worse than dead but the conceptual and practical bases for such states is highly controversial and I suspect, especially from a policy point of view, that nothing much is lost by ignoring them.²⁷ It is more important to have a comparable endpoint for the top of the scale. This does not have to be full health – in fact it would be better if it was 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.²⁸ For example, those reporting pain in the British Household Panel Survey might compare their subjective well-being to others in pain, which would lead to over-interpretation of the degree of any adaptation. In other words, respondents who are in pain might overstate their levels of well-being as they are comparing themselves with others in pain rather than the population as a whole. It is possible that responses may shift more for some conditions than for others and this would bias any QALY estimates. To address this problem, respondents in a recent study by the World Health Organisation were asked to rate standardised vignettes describing other people's health alongside their own health satisfaction, so that the latter could be recalibrated

²⁷ That 83 of the 243 health states defined by the EQ-5D (for an explanation of EQ-5D see footnote 8 above) are considered to be worse than death when using the 10-year TTO method is another example of the problems with using preferences. Having said this, most of these are very severe states that thankfully very few people experience, and so they would

be unlikely to have much effect on the QALY results for most technologies submitted to NICE.

²⁸ See: Sprangers M.A.G. and Schwartz, C.E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science and Medicine*, 48, 11, 1507-1515.

against the former.²⁹ This approach led to self-reported health declining more rapidly with age, once each individual's health rating had been recalibrated against her rating of the vignette, as it accounted for older people comparing their health to other old people, as opposed to the population in general.

However, response shift cannot explain all changes in preferences that take place: there is evidence of adaptation when physiological measures are used, which should be less prone to response shift. There is behavioural evidence, for instance, that war veterans with more severe past injuries could hold their finger in hot water for longer than veterans with less severe past injuries before classifying it as painful.³⁰ This is an example of a physiological adaptation to a health state.

There are more fundamental concerns about the nature of adaptation itself. All else being 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.³¹ In light of any costs associated with adaptation, it may be unjust if some patients lost competitive advantage in the race for scarce resources because their effort diminished the estimated value of treatments for them as compared to other patients who did not expend the same effort. Thus, in the case of effort-driven adaptation, resource allocation may require weighting such that some patients are not exposed to both the costs of illness and the loss of resources.

²⁹ Salomon, J.A., Tandon, A. and Murray, C.J.L. (2004). Comparability of self rated health: cross-sectional multi-country survey study using anchoring vignettes. *British Medical Journal*, 328 (7434), 258-61.

³⁰ Dar, R., Ariely, D. and Frenk, H. (1995). The effect of past injury on pain threshold and tolerance. *Pain*, 60, 189-93.

³¹ The opposite is true, of course, for life-saving or life-extending interventions, where complete adaptation to a health state would mean that the number of QALYs gained from saving someone's life in that state would, all else equal, be the same as the number of QALYs gained from saving the life of someone in full health.

On the other hand, if this argument is accepted, then an advantage gets created for those who have adapted. Treatments for such people will get greater priority than is 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. The fact that we appear not to adapt to mental health problems (perhaps by their very nature) strengthens the case for policy interventions in these areas.

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 reflect the former interpretation. In developing a 'reference case' for the economic appraisal of new health technologies, NICE has followed this tradition and recommends the use of preference-based valuations of health.³² Further, they recommend that those preferences be elicited from members of the general public. The choices that respondents make in valuation studies should be guided by accurate assessments of the well-being associated with the consequences of those choices. 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 not make us happier and vice versa.³³

I was closely involved in the study that elicited time trade-off valuations for health states from the general public, the results of which 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' and so, all else being equal, the benefits from treating the former would be greater and so would be given greater priority by NICE. It is my strong intuition (and the evidence used to compile Table 1 bears this out) that allocating resources in this way would not improve well-being – as it is felt and experienced – as much as it could. I have come to consider that the

³² It is worth saying that NICE's reference case will not generate one unique set of QALY values. There are considerable degrees of freedom involved in the choice of classification system and valuation method, both of which have been shown to have significant effects on resulting QALY weights.

³³ Gilbert, D. T. and Wilson, T. D. (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.

use of values like those I was involved in generating potentially 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. I think that the problems with focusing effects – as Kahneman has said on many occasions, 'nothing in life is as important as when you are thinking about it' – are pervasive and call into question most preference-based valuations. Of course, preference-based methods may be useful in other contexts, such as predicting behaviour, but I contend that they are of little use for valuing health. This is quite apart from all the biases and heuristics that are inherent in the valuations, which I have not even touched on.

It is my contention that an individual's assessment of her life may offer a more defensible measure of well-being than the satisfaction of her actual preferences, and that measures of subjective well-being 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, subjective well-being 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 which is being valued, therefore removing the risk of strategic behaviour, or artificially created preferences.

Whilst the onus of responsibility for the production of 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 British Household Panel Survey, which could be explored further to illustrate the effects of different health conditions on subjective well-

being. However, we also require large longitudinal studies that look at subjective well-being pre- and post-intervention to determine the benefits from intervention.

This is all for the future and you might like to know what I would do right now. Well, if I was asked to reallocate healthcare resources on the basis of what we currently know, I would probably put more money into mental health services and take it out of treatments for improving physical functioning but, as any academic would always say, more research is needed. I hope I have encouraged others to join me in a research endeavour that seeks to ration healthcare in ways that improve the real experiences of how people think and feel about their lives rather than in ways that satisfy their hypothetical preferences over how they imagine thinking and feeling.

ABOUT THE AUTHOR

Paul Dolan is Professor of Economics at Tanaka Business School, Imperial College London. The general theme of his research activities is how individual and social well-being should be defined, measured and distributed for the purposes of informing public policy. He was previously Founder-Director of the Centre for Well-being in Public Policy at the University of Sheffield and a Visiting Research Scholar at Princeton University. Paul was awarded a Philip Leverhulme Prize in Economics in 2002 for his contribution to health economics.

ACKNOWLEDGEMENTS

This paper draws heavily on ideas developed with Daniel Kahneman when I was a Visiting Research Scholar at Princeton University. I am also gratefully indebted to Rob Metcalfe and Mat White for comments on drafts of this paper, and to Tessa Peasgood for the data analysis and for general advice and argument. In addition to their own comments on the paper, I am indebted to Barney Gough at the SMF for getting to the bottom of what NICE requires in its technology assessments and to Ann Rossiter for having the foresight to suggest this paper in the first place. All of these people have directly contributed towards my own well-being to a much greater extent than I would have forecast. I take full responsibility for the views expressed in this paper.

£5