Work In Progress – Please do not quote without authors' permission

'Naming and framing': an investigation of the effect of disease labels on health state valuations

Suzanne Robinson

Stirling Bryan

Health Economics Facility
Health Services Management Centre
Park House
40 Edgbaston Park Road
Birmingham B15 2RT

Tel: 0121 414 3011
Fax: 0121 414 7051
Email: smr910@bham.ac.uk

Paper presented at the Health Economists' Study Group Meeting, January 2001, University of Oxford

1 Introduction

Economic analyses are increasingly being employed in formal resource allocation decision-making processes in health care. For example, in the UK the National Institute for Clinical Excellence (NICE) requires economic evidence to be provided as part of submissions from industry and other interested parties; and this is also seen in similar circumstances in both Australia and Canada (Rawlins, 1999). The consequence is that the methods being employed by economic analysts are increasingly subject to close scrutiny. Clearly this represents both an appropriate and a positive development. The implication is that in areas where a variety of alternative methods are currently being employed by health economists, as a minimum some explanation for the variance is required, and at best a preferred position should be identified. One such area of methodology concerns the elicitation of utility scores on hypothetical health or disease states for use in the construction of quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs). This area represents the focus for this paper.

A number of research issues remain unresolved in the area of health state valuation. These might be grouped under four general headings:

- The description of health states
- The valuation method to use
- Data collection mechanisms
- The general framing of questions

There is no gold standard with regards to the descriptive system that should be used when measuring quality of life in the context of economic analyses. However, Gerard *et al* (1993) provide an indication of the desirable attributes of such a system:

> "valid health descriptions, must be easily understood, comprehensive yet sensitive, reliable and relevant. If health descriptions were to be judged invalid, little confidence could be placed in the subsequent quality of life values elicited".

The work of the EuroQol Group, in particular, has now provided us with an off-the-shelf health state descriptive system that satisfies many of the criteria laid down by Gerard *et al* (Brooks, 1996). In addition, other generic descriptive systems to define health states are also available, for example, the Health Utilities Index (Feeny *et al*, 1995) and the 15D (Sintonen, 1981).

A number of alternative approaches to the valuation of health states are currently being used, including visual analogue scales (VAS), time trade-off (TTO) techniques, standard gamble (SG) methods, and equivalence of numbers approaches, also referred to as person trade-off (PTO). The strengths and weaknesses of each approach largely concern:

- the practical (i.e. ease of administration of the exercise),

- the conceptual (i.e. whether the utility score elicited using the technique represents a measure of societal value), and
- the theoretical (i.e. the theoretical basis of the exercise).

These issues are discussed in detail elsewhere (for example, see Brazier *et al*, 1999).

Unanswered questions relating to the appropriate data collection mechanism to employ include:

- Should data be collected via face-to-face interviews or can a postal survey deliver reliable and valid data?
- What degree of consideration and deliberation should respondents be encouraged to engage in prior to providing responses?

A host of more general unresolved framing issues also exist. For example, what time horizon in the health states in question should respondents be asked to consider, and to what extent should respondents be provided with supporting information on the degree of adaptation patients typically make when faced with the challenges presented by poor health states? As part of this general framing category is the issue of the labelling of health states with a description of the disease to which the health state description refers. This issue is the particular focus of this paper.

The topic of disease labelling is important and of current relevance given the ongoing debate concerning the use of alternative summary measures of health: the QALY and the DALY (Murray & Lopez, 2000; Williams, 2000a; Williams, 2000b). One of the central differences between the two measures concerns the presentation of health states. In a QALY-framework, health states are presented without a disease label since the objective is to use the valuations obtained as generic values applicable across a wide range of possible diseases. In contrast, within a DALY-framework, the explicit purpose in collecting data on values is to facilitate the estimation of the burden of specific diseases and so typically the disease label, along with a description of the condition, are provided to respondents in addition to the health state description.

The importance of providing labels in health state valuation exercises has been investigated by a number of researchers (for example, Gerard *et al*, 1993; Rabin *et al*, 1993; Sackett & Torrance, 1978; O'Connor, 1989; Smith & Dobson, 1993). The study by Rabin *et al* (1993) looked at the impact on utility scores of assigning disease labels to health states. Their results suggest that when 'mental conditions' were explicitly labelled, such as mental handicap, schizophrenia and dementia, then lower utility values tended to be assigned. They conclude that explicit reference to clinical diagnosis does affect the responses individuals give when asked preference elicitation questions. In contrast, the study reported by Gerard *et al* (1993), which considered issues relating to the presentation of breast cancer health descriptions, found valuations not to be sensitive to general framing and labelling factors. However, they also report that significant differences in valuations were found when descriptions were written in the third party and when the term "cancer" was used.

This paper reports the results of a pilot study which sought to investigate the impact of disease labels on health state values across a wide range of conditions and using three health state valuation techniques.

2 Methods

2.1 Overview

The principal objective of this research was to compare two sets of health state valuations: a set obtained when states were framed using disease labels and descriptions, and a set derived from data collection exercises where generic health state descriptions were used. A secondary objective was to explore the extent to which the effect of labelling differed across alternative approaches to the elicitation of values.

This research exercise was conducted as an adjunct to the European Disability Weights project and so used methods developed as part of that project. The EDW project involves researchers (public health, epidemiology and health economics) from seven countries over the period 1998-2001. Full results of the project are to be presented at the European Public Health Association conference in Paris, December 2000. One objective of this project is to construct a common set of European disability weights for use in estimating DALYs. The empirical work, debates and discussion that occurred during the development of methods used in the project will be presented elsewhere.[1]

2.2 Scenarios presented

Disease-specific scenarios (disease labels)

The study used 9 scenarios that detailed the disease and disease stage. These scenarios were exactly the same as those used in the EDW project, and related to the following diseases:

- Breast cancer
- Vision disorder
- Diabetes
- Acute myocardial infarction
- Quadriplegia
- Low back pain
- Depression
- Asthma
- Stroke

[1] The EDW Project will be publishing two reports in 2001.

An example of a disease-specific scenario used is given in Box 1. The scenarios comprised the disease label and a generic description of health. This description is clearly based on the EuroQol EQ-5D instrument but is different in that it additionally includes a cognitive component. This modified version of the EQ-5D was used in order to maintain consistency with the EDW project in which the scenarios were tested and found acceptable by both lay and medical survey participants. As shown in Box 1, the 'disease label' used in this study was not just the title or name of the disease to be valued, but the label also contained a sentence describing the health state in greater detail.

Generic scenarios (no disease label)

The generic scenarios were produced by removing the disease label from the scenario description, leaving only the health state description (see Box 2). Therefore the only real difference between the two sets of scenarios was the disease label/disease description. The same set of 9 health states descriptors was used in both the disease specific and generic scenarios.

2.3 Methods of health state valuation

The process of health state valuation was similar to the protocol as used in the EDW project. All participants were asked to value the various health states using three valuation approaches: VAS, PTO and TTO.

- *VAS method*
  The VAS required individuals to value health states on a scale ranging from 0 (worst imaginable health state) to 100 (best imaginable health state).

- *PTO method*
  The PTO method attempts to estimate the social or societal value of different health states. Patrick *et al* (1973) who referred to it as the equivalence of numbers procedure originally developed this approach and it has been developed further by Nord (2000), who refers to the approach as person trade-off (PTO). The EDW Group developed the version used in this study (referred to as PTO3), with advice from of Erik Nord (for further details see Nord, 2000). The format and content of the PTO3 method are essentially as follows:

Imagine that you are a decision maker. You have a choice between two programs that will reduce the incidence of disease in a few years from now.

- *Program A will prevent the occurrence of a rapidly fatal disease in 100 people in your country.*

- *Program B will prevent the occurrence of disease X (chronic state described in detail) in N people in your country.*

The programs are in all other respects equal.

Choose the value for N that would make you indifferent between the two programs. When answering, please disregard possible economic aspects.

(reproduced from Nord, 2000)

In line with the EDW project, the PTO method employed in this study had two stages. In stage one, a single comparison was made between a programme to prevent fatal disease and a programme to prevent quadriplegia. In stage two, a series of 8 comparisons were made, each involving a programme to prevent quadriplegia and a programme to prevent one of the further 8 diseases. The implication of this is that the PTO response for quadriplegia is not comparable with the PTO responses for the other conditions, since the comparator is different. It should also be noted that when PTO was used *without* disease labels, the label of quadriplegia for the comparator programme was not provided to respondents.

- *TTO method*
  The TTO method focuses on the individuals own values of different health states. The participant is asked to choose between two alternatives: years in full health or years in the health state being valued. The participant is given the opportunity to trade-off years of healthy life to move from the health state described to 'full health'. The TTO method used in this study considered a fixed duration of 10 years, in order to maintain consistency with the EDW project.

2.4 Sample and data collection

This pilot study was conducted on a convenience sample of 26 Masters students, taking a one week course in Health Economics at the University of Birmingham. Participants were split into two groups: one group were asked to value the disease specific scenarios and the other group was asked to value the generic scenarios. The data collection exercises for the groups were undertaken independently in different locations and so there was no opportunity for discussion between groups prior to providing responses. Each group had a single facilitator.

The first stage of the data collection session involved the health states being presented to the subjects, and this was done by the group facilitator describing all 9 scenarios and providing each participant with 9 show-cards containing the health state scenarios (either with or without the disease labels). Each valuation method was then administered.

- Participants were first asked to place the 9 health scenario show-cards onto the VAS.

- The PTO method was the second method to be presented. It was introduced using an example of severe vision disorder (see Boxes 1&2). Subjects were then asked individually to value the 9 scenarios using the PTO method, using the approach described above.

- The TTO method was then used and was also introduced using the example of severe vision disorder. Participants were then asked to value the 9 remaining health states using the TTO method.

Finally, subjects were asked to check their response sheets to make sure they were 'satisfied' with the valuations they had given.

2.5 Data Analysis

Frequency distributions of responses were plotted for all diseases and all methods, allowing an initial visual inspection of the data. The statistical significance of differences between labelled and non-labelled scenarios was explored using the Wilcoxon test (Altman, 1991). These comparisons were made separately for each valuation method.

3 Results

Of the 26 respondents in the pilot study, 17 were female. The mean age of the whole group was 32 years (range: 21-43). Twenty-six respondents had a first degree and 3 stated that they had some previous experience in health state valuation. The two groups were well matched in terms of the age distribution ('no label' group mean: 35.5 years; 'with label' group mean: 32.2 years), although there was some imbalance in the sex distribution: 6 respondents in the 'no label' group were female whilst 11 respondents in the 'with label' group were female.

The maximum possible number of data points, in terms of responses to health state valuation choices was 702; the total number provided was 631. A single respondent from each group refused to undertake both the PTO and TTO exercises, but did complete the VAS.

Summary statistics relating to both 'with label' and 'without label' responses on VAS, TTO and PTO are reported in Tables 1, 2 and 3, respectively. The difference between the 'with label' and 'without label' median responses for each disease and for each of the three methods is shown graphically in Figures 1, 2 and 3. Figure 1 indicates that for 6 (of the 9) health state scenarios the effect of the label was for a lower VAS score to be given. The difference between 'with label' and 'without label' VAS scores was statistically significant (at the 5% level) for two scenarios: stroke ($p=0.036$) and quadriplegia ($p=0.016$). Both of these scenarios had higher VAS scores when no label was used.

A similar picture emerges from the TTO data (see Figure 2): when health state scenarios were described using a disease label the median respondent was willing to give up a larger number of life-years in order to avoid the health state. This was the case for all health state scenarios, except for low back pain where the 'with label' and 'without label' medians were the same. The interpretation of these results is that health state scenarios were consistently viewed as being poorer when the disease label was included. It should,

however, be noted that none of the differences between medians reached conventional levels of statistical significance.

The PTO data (see Table 3 and Figure 3) appear to reveal findings that are at odds with those indicated by the VAS and TTO data. For all health state scenarios, except breast cancer and diabetes, the median PTO response was higher for the situation where a disease label was used. The interpretation of this finding is that scenarios where no label was used were viewed as being more severe than the same scenario where a disease label was provided. The differences between medians were statistically significant (at the 5% level) for two diseases: depression ($p=0.012$) and low back pain ($p=0.031$). However, these PTO results should be treated with caution because the comparator programme was different for the 'with label' exercise (i.e. defined as 'quadriplegia') and the 'without label' exercise (i.e. not defined as 'quadriplegia' but simply described using the modified EQ-5D descriptor). The results from both the VAS and TTO data suggest that respondents viewed the health state scenario differently when the label of 'quadriplegia' was used.

4 Discussion

The work reported in this paper represents a pilot study and so one should be very cautious in drawing conclusions solely on the basis of the data presented here. The general finding, from the VAS and TTO results in particular, is that assigning labels to health states appears to have an effect: when a label was used the health scenarios tended to be viewed as more severe conditions.

The results presented in this paper are broadly consistent with the findings reported by others, although it is important to note that the methods employed by other researchers have varied widely. For example, in the study by Sackett & Torrance (1978) they presented members of the general public with three pairs of health state descriptions, each pair containing one description with a label and one without a label. They found that the label did have significant effects on the utility values obtained, and that these label-induced variations went in both directions: 'tuberculosis' was given a higher value than 'unnamed contagious disease', whereas 'mastectomy for breast cancer' was given a lower value than 'mastectomy for injury'. This finding is in line with the work of Gerard *et al* (1993) and Rabin *et al* (1993) who both also noted the negative impact on valuations of the term 'cancer'. The possibility for utility values to be systematically higher or lower as a result of the use of a disease label was also the finding of Rabin *et al* (1993). They showed that where the label referred to mental conditions, such as schizophrenia, the values tended to be consistently lower; and where physical conditions were being considered, such as osteoarthritis or burns, the effect of the label was to induce a higher valuation. The findings reported in this paper are not fully supportive of the results of Rabin *et al* (1993): our data suggest that the label tends to be associated with lower health state valuations across both mental and physical conditions.

One of the strengths of the study reported here is that it was conducted alongside a larger study (i.e. the EDW project) and so was able to build upon the work undertaken as part of that project. For example, the health state scenarios and descriptions used here had previously been tested on both medical and lay participants. However, the use of materials from another study also imposes restrictions. This is clearly seen in the use of PTO3 which was not directly suitable for use in a study looking at the effects of disease labelling because of the fact that a medical condition (i.e. quadriplegia) is used as the comparator programme in the choices presented. The PTO method was included here since it was used by the World Health Organisation in the Global Burden of Disease Study in 1996. The developers of the PTO3 method claim that it overcomes some of the criticisms of earlier versions relating to ethical issues and concerns about comprehensibility (see Nord, 2000; Arnesen & Nord, 1999).

Recent debate in the literature has indicated the importance of discussion and deliberation in the process of eliciting preferences for use in resource allocation decisions (see Dolan *et al*, 1999). It is possible different results may have been obtained if discussions had been encouraged as part of the group sessions, especially in the case where disease labels were added because of the extra information that this provides. However, practical constraints did not allow this to be the case in this study. It should be noted that this represents a departure from the protocol used in the EDW study.

As indicated above, this paper describes the pilot phase of this work and data collection on a larger sample of respondents, using a combination of quantitative and qualitative methods, is currently in the planning phase. This follows, in part, from the recommendations of Roberts *et al* (1999) for further qualitative work in this field to understand further how clinical information is interpreted by respondents and the effect this has on the response to preference elicitation questions. However, if the finding of this pilot holds more generally then the issue of labelling of health states has to be confronted. In essence the issue concerns whether it is appropriate for health care resource allocation decisions to be based, in part, on preferences (and/or prejudices) relating to diseases. That is, should individuals be advantaged (or disadvantaged) in accessing health care because of the disease they have, all other things being held constant?

Box 1: Example disease-specific scenario

Vision Disorder	SVIS
Mild /moderate vision **Severe vision disorder** ←	

Patient is unable to read small newspaper print and has great difficulty or is unable to recognise faces at 4 meters distance

No problems in walking about
- Some problems with washing or dressing self
- Some problems with performing usual activities (e.g. work, study, housework, family or leisure activities)

No pain or discomfort
- Moderately anxious or depressed

No problems in cognitive functioning (e.g. memory, learning ability, concentration, comprehension)

Box 2: Example generic scenario

Health state B

No problems in walking about
- Some problems with washing or dressing self
- Some problems with performing usual activities (e.g. work, study, housework, family or leisure activities)

No pain or discomfort
- Moderately anxious or depressed

No problems in cognitive functioning (e.g. memory, learning ability, concentration, comprehension)

Table 1: Summary statistics for VAS responses

	Label		*No label*	
	Mean (SD)	*Median (IQR)*	*Mean (SD)*	*Median (IQR)*
Quadriplegia	7.91 (8.81)	5 (2-9)	23.46 (21.98)	15 (8-34)
Depression	26 (19.07)	22 (9-38)	25.76 (18.47)	20 (14-25)
Low back pain	62.61 (18.82)	66 (50-76)	57.07 (14.98)	62 (50-68)
Stroke	24 (13.98)	24 (16-28)	37.75 (15.19)	40 (25-50)
Asthma	52.61 (17.82)	60 (38-66)	53.07 (14.94)	58 (46-60)
Vision disorder	74.61 (24.87)	76 (70-92)	71.84 (12.71)	84 (76-84)
Breast cancer	62.16 (25.99)	64 (49-82)	71.30 (13.72)	74 (70-80)
Diabetes	57.25 (14.21)	55 (50-70)	66.53 (14.25)	68 (64-78)
AMI	32.61 (12.50)	38 (26-40)	36.69 (20.13)	40 (27-50)

Table 2: Summary statistics for TTO responses

	Label		*No label*	
	Mean (SD)	*Median (IQR)*	*Mean (SD)*	*Median (IQR)*
Quadriplegia	3.08 (3.17)	2 (1-4.5)	5.81 (3.60)	7 (2-10)
Depression	5.16 (2.85)	5.5 (2.5-7.5)	6.27 (3.77)	6 (4-10)
Low back pain	8.91 (2.84)	10 (9-10)	8.2 (3.15)	10 (7-10)
Stroke	6 (2.41)	6 (5-8)	7.2 (2.9)	8 (5-10)
Asthma	7.41 (2.74)	8.5 (6.5-9)	8.2 (2.93)	10 (7-10)
Vision disorder	8.25 (2.98)	9.5 (7.5-10)	9 (2.82)	10 (10-10)
Breast cancer	8 (3.01)	9 (8-10)	8.7 (2.83)	10 (8-10)
Diabetes	8.08 (2.71)	9 (8-9.5)	8.5 (2.83)	10 (8-10)
AMI	6.75 (2.59)	7.5 (5.5-8)	6.9 (3.34)	8 (3-10)

Table 3: Summary statistics for PTO responses

	PTO - Label	PTO - No label
	Median (IQR)	Median (IQR)
Quadriplegia[1]	155 (75 - 200)	150 (75 - 500)
Depression[2]	8,500 (200 - 150,000)	175 (75 - 1,000)
Low back pain[2]	25,000,000 (200,000 - 60,000,000)	5,000 (1,000 - 1,000,000)
Stroke[2]	6,250 (750 - 95,000)	1,000 (400 - 1,000,000)
Asthma[2]	75,000 (3,750 - 35,000,000)	2,000 (750 - 50,000)
Vision disorder[2]	60,000,000 (300,000 - 60,000,000)	10,000,000 (4,000 - 60,000,000)
Breast cancer[2]	10,000 (5,000 - 1,000,000)	50,000 (1,000 - 60,000,000)
Diabetes[2]	12,500 (5,000 - 30,000,000)	25,000 (1,500 - 60,000,000)
AMI[2]	4,000 (1,000 - 15,000,000)	1,000 (500 - 5,000)

NB: Means not reported since distributions highly skewed

[1] Comparator programme involved fatal disease

[2] Comparator programme involved quadriplegia

Figure 1: Comparison of 'with label' and 'without label' median VAS valuations

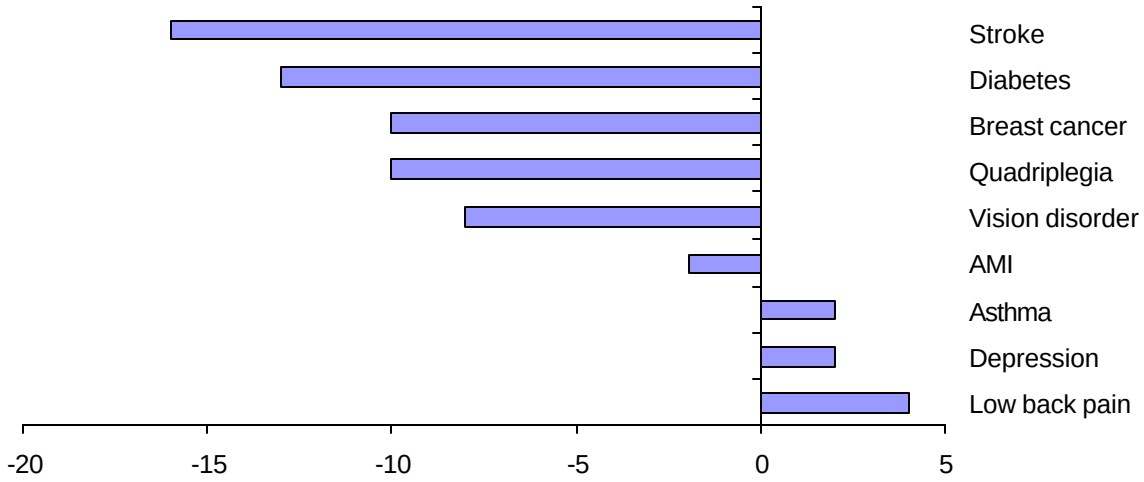
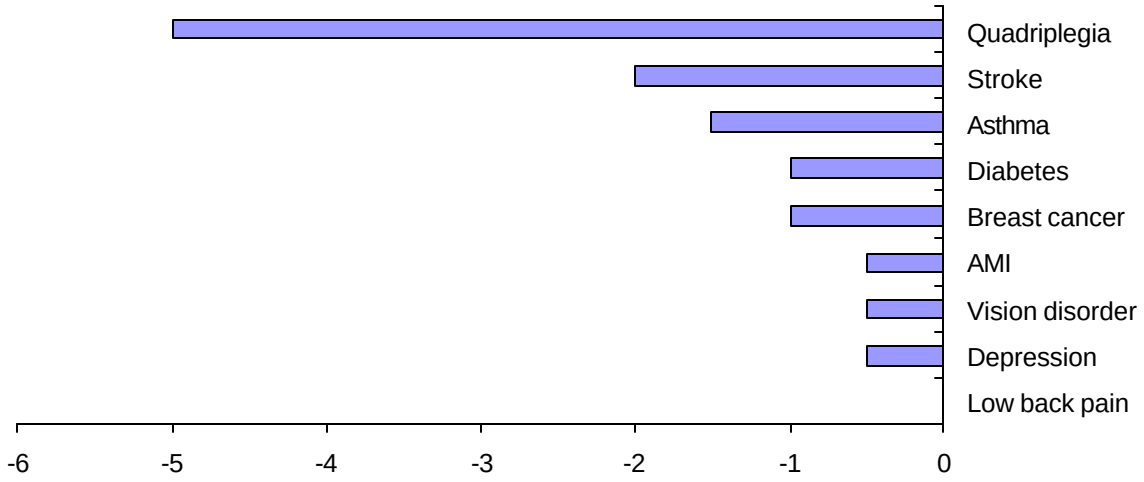
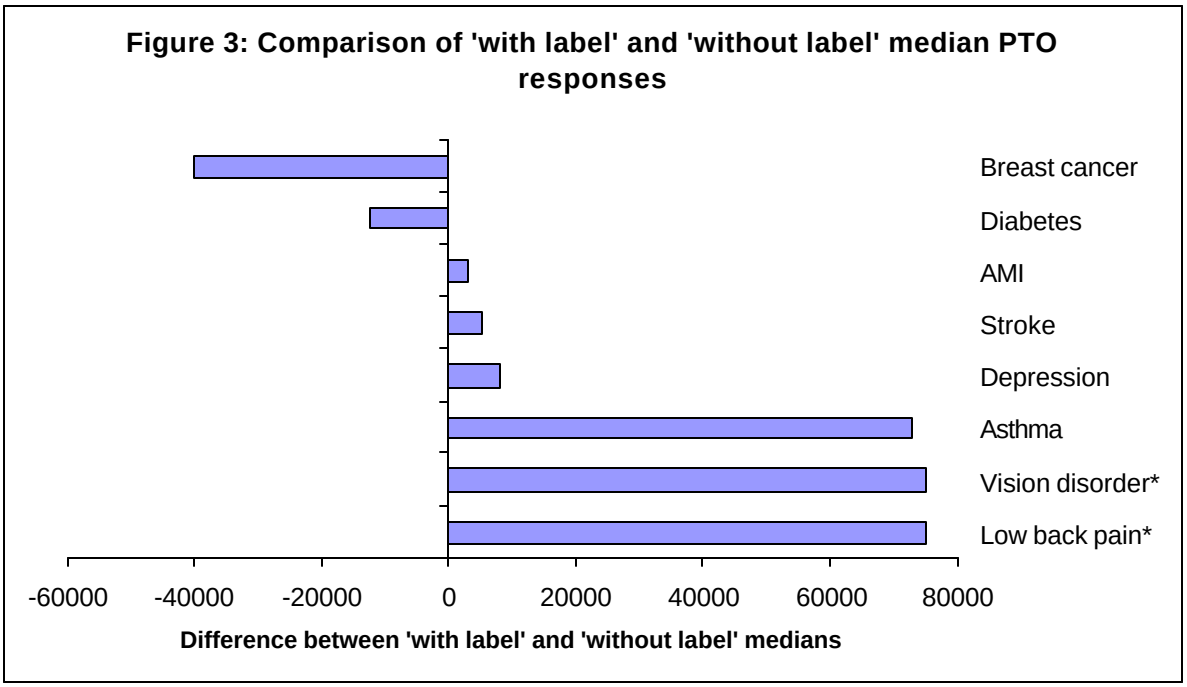


Figure 2: Comparison of 'with label' and 'without label' median TTO responses





* Bar truncated at 75,000. Actual difference between medians for low back pain: 50,000,000; and for vision disorder: 24,995,000.

References

Altman DG Practical statistics for medical research. Chapman & Hall London (1991).

Arnesen T and Nord E .The value of DALY life: problems with ethics and validity of disability adjusted life years. BMJ 1999;319:1423-1425.

Bettman and Kakkar R, Effects of information presentation format and consumer information acuisition strategies. Journal of Consumer Research 1977;3:233.

Bettman JR and Sujan M. Effects of framing on evaluation of comparable and noncomparable alternatives by expert and novice consumers. Journal of Consumer Research 1987;14:141.

Biehal G and Chakravarti D. Consumers' use of memory and external information in choice. Journal of Consumer Research 1986;12:382.

Brazier J and Dixon S, The use of condition specific outcome measures in economic appraisal. Health Economics 1995;4:255.

Brazier J, Deverill M, Green C, Harper R, and Booth A. A review of the use of health status measures in economic evaluations. Health Technology Assessment 1999;3[9].

Brooks R, EuroQol: The Current State of Play. Health Policy 1996;37:53.

Dolan P, Cookson R, and Ferguson B. Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. BMJ 1999;318:916.

Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems, Health Utilities Index. Pharmacoeconomics 1995;7:490-502.

Fischhoff B, Value Elicitation. American Psychologist 1991;46:835.

Gerard K, Dobson M, and Hall K. Framing and labelling effects in health descriptions: Quality adjusted life years for treatment of breast cancer. Journal of Clinical Epidemiology 1993;46:77.

Johnston K, Brown J, Gerard K, O'Hanlon M, and Morton A. Valuing temporary and chronic health states associated with breast screening. Social Science and Medicine 1998;47:213.

Kirsch J and McGuire A. Establishing health state valuations for disease specific states: An example from Heart Disease. Health Economics 2000;9:149.

Llewellyn HT, Sutherland HJ, Tibshirani R, Clampi A, Till JE, and Boyd NF. Describing Health States. Medical Care 2000;22:543.

Murray CJL and Lopaz AD. Progress and directions in refining the global burden of disease approach. Health Economics 2000;9:69.

Nord E. My goodness-and yours. (http://www.eriknord.no/health/Dalyhistory.htm), 2000.

O'Conner A. Effects of framing and level of probability on patients' preferences for cancer chemotherapy. Journal of Clinical Epidemiology 1989;42:119.

Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health Services Research 1973;8:228-45.

Payne JW and Bettman JR. Behavioural decision research: A constructive processing perspective. Annual Review of Psychology 1992;43:87.

Rabin R, Rosser RM, and Butler C. Impact of diagnosis on utilities assigned to states of illness. Journal of the Royal Society of Medicine 1993;86:444.

Rawlins M. In pursuit of quality: the National Institute for Clinical Excellence, The Lancet 1999;353:1079.

Roberts T, Bryan S, Heginbotham C, and McCallum A. Public involvement in health care priority setting. Health Expectations 1999;2:235.

Sintonen H. An approach to measuring and valuing health states. Soc Sci Med 1981;15C:55-65.

Smith R and Dobson M. Measuring utility values for QALYs two methodological issues. Health Economics 1993;2:349.

Stouthard M, Essink-Bot ML, Bonsel G, Barendregt JJ, Kramer PGN, van de Water HPA, Gunning-Schepers LJ, and van der Maas PJ. "Disability weights for diseases in the Netherlands". Department of Public Health Erasmus University Rotterdam, 1997.

Tversky A and Kahneman D. The Framing of Decisions and the Psychology of choice. Science 1981;211:453.

Williams A. Calculating the global burden of disease. Health Economics 2000;8:1.

Williams A. Comments on the response by Murray and Lopez. Health Economics 2000;9:83.

Acknowledgements

We would like to thank members of the EDW Project Team with whom we have worked, who have provided us with the inspiration for this study and the opportunity to take this work forward. In addition, particular thanks goes to our colleagues at HSMC, notably Lisa Gold for her support with this work and for her comments on an earlier draft of the paper, and Tracy Roberts for her support in data collection.