

# A cardiovascular extension of the Health Measurement Questionnaire

Andrew J Martin, Paul P Glasziou, R John Simes

## Abstract

**Objective**—To investigate the psychometric properties of a cardiovascular extension of an existing utility-based quality of life questionnaire (Health Measurement Questionnaire). The new instrument has been named the Utility Based Quality of life—Heart questionnaire, or UBQ-H.

**Design**—Explored the test-retest reliability, construct validity, and responsiveness of the UBQ-H.

**Patients**—A sample of 322 patients attending cardiac outpatient clinics were recruited from two large metropolitan teaching hospitals. A second sample of 1112 patients taking part in the LIPID trial was also used to investigate the validity and responsiveness of the UBQ-H.

**Results**—Ninety per cent of all UBQ-H questionnaires were returned, and item completion rates were high (median of less than 1% missing or N/A answers). Cronbach's  $\alpha$  measure of internal consistency for the scales ranged between 0.79–0.91, and each item was also most strongly correlated with its hypothesised domain than alternative domains. The intra-class test-retest reliability of the UBQ-H scales ranged from 0.65 to 0.81 for patients with stable health. Results supported the construct validity of the UBQ-H. The UBQ-H was significantly correlated with other information on quality of life (for example, General Health Questionnaire) as anticipated. The instrument was able to distinguish between contrasted groups of patients (for example, with versus without symptoms of dyspnoea, prior myocardial infarction versus none, etc), and was responsive to changes in health associated with adverse events requiring hospitalisation.

**Conclusions**—The modifications made to the Health Measurement Questionnaire has resulted in an assessment designed for cardiovascular patients that has proved to be both reliable and valid.

(J Epidemiol Community Health 1999;53:548–557)

NHMRC Clinical Trials Centre,  
University of Sydney,  
Sydney, Locked Bag 77  
Camperdown, NSW  
2050, Australia  
A J Martin  
R J Simes

Department of Social and Preventive Medicine, University of Queensland,  
Herston, Australia  
P P Glasziou

Correspondence to:  
Mr A J Martin.

Accepted for publication  
30 October 1998

Quality of life instruments provide a valuable means of assessing treatment effects and the impact of disease across numerous aspects of patients' lives, but often fall short of providing a summary utility index for overall quality of life. Such an index can be useful for exploring the impact of trade offs between aspects of quality of life, and is needed for the calculation of quality adjusted survival (for example, QALYs).<sup>1</sup> Quality adjusted survival can be a valuable measure

for summarising treatment outcomes across both quality of life and survival end points, and is also used in a type of economic evaluation known as a cost utility analysis.

Several years ago a multidimensional quality of life assessment that could also provide a utility score was sought for use in a secondary prevention trial with post-myocardial infarction and unstable angina patients (Long term Intervention of Pravastatin in Ischaemic Disease—LIPID trial<sup>2</sup>). The Health Measurement Questionnaire<sup>3</sup> showed potential for use in the LIPID trial because it had been successfully used with cardiovascular patients in a trial of thrombolytic therapy for myocardial infarction,<sup>4</sup> it was self administered, and the scoring algorithm was based on Rosser's Classification Index,<sup>5,6</sup> which has been applied in a variety of clinical studies.<sup>7–9</sup> None the less, the Health Measurement Questionnaire was not specifically designed for use with cardiovascular patients. Data from the thrombolytic study suggested a number of items were less relevant for cardiovascular patients with few patients reporting any problems on these.<sup>4</sup> Furthermore the instrument was lacking clinically important items on physical activities and overall quality of life.

A research programme was consequently undertaken to develop a version of the Health Measurement Questionnaire for use in the LIPID trial. The Utility Based Quality of life—Heart, or UBQ-H, questionnaire was developed from the Health Measurement Questionnaire in three main steps. Firstly, the less relevant items (for example, hearing, vision, writing, speaking, and incontinence) identified in the thrombolytic trial were discarded.<sup>4</sup> Secondly, the response format of some items was modified. Visual analogue scales (a blank line 10 cm long) were changed to numbered rating scales based on the recommendation of Moinpour and colleagues.<sup>10</sup> The dichotomous "yes/no" response options for items on ability to engage in various activities were changed to four order categories ranging from "no limitation" to "unable to do activity at all". Thirdly, items traditionally associated with cardiovascular disease were added. This included items on physical ability (for example, ability to climb stairs) and specific symptoms (for example, muscle aches and pains). Three measures of overall quality of life were also included. These comprised a time trade off question, a rating scale, and a simple ordinal self rated health item. The self administered time trade off item is based on that successfully used with post-myocardial infarction patients in the thrombolytic study.<sup>4</sup> The item has been

If you consider yourself in less than full health, please help us determine how important quality of life versus length of life is to you by answering the following hypothetical question.

Imagine a friend who is expected to live for 15 years with the same quality of life as you have now. Suppose treatment could restore them to full health, but would shorten their life. At most, how much time would you advise giving up out of 15 years?

I would advise giving up at most \_\_\_\_\_ years and/or \_\_\_\_\_ months in order to return to full health.

Figure 1 Self administered time trade off question.

reproduced in figure 1. The rating scale is of similar origin to that used in the EuroQoL questionnaire<sup>11</sup> but was anchored by full health and death. Versions of the ordinal self rated health item (excellent, good, fair, poor) have been used in several other instruments, including the SF-36.<sup>12</sup> Table 1 lists the UBQ-H items and indicates the range of response options for each item. The items are grouped in the table according to the quality of life domain that they are hypothesised to sample. Four quality of life domains are represented: physical ability, psychological distress, social/usual activities, and self care.

The aim of this paper was to investigate the psychometric properties of the UBQ-H. The Rosser Index can still be applied to the UBQ-H to provide summary measures of physical disability and psychological distress and to estimate a utility score. However, this scoring approach ignores information from the newly added items. Furthermore, as the rating method used to devise the Rosser Index was magnitude estimation rather than a standard gamble (or equivalent), the scores it produces may not have true utility properties.<sup>6</sup> A future goal was to therefore develop a more responsive scoring system calibrated against cardiovascular patients' self assessed utilities. The psychometric investigation detailed herein was an important initial step in this research goal.

## Methods

A sample size of at least 200 cardiovascular patients was sought from two large metropolitan teaching hospitals. Patients from the first hospital were sampled from the population attending the cardiac outpatient clinic. Patients in poorer health were sought from the heart transplant and heart failure clinic run at the second hospital. The sample size target of 200 was chosen to provide reasonable precision in estimating intraclass test-retest reliability (for example, the 95% confidence intervals on an estimate of reliability with a sample of 200 would be  $\pm 0.15$ ), and to provide sufficient power for the validation analyses (for example,  $n=200$  gives 80% power to detect an effect of around 0.4 of a standard deviation with 95% confidence). The eligibility criteria were: aged between 24 and 75 years (to match LIPID patients' characteristics), physically able to meet the requirements of the study, sufficient English, and provision of informed consent.

### STUDY DESIGN

Information on patients' health/quality of life was collected using the UBQ-H and the following measures.

### Physical functioning

A modification of the New York Heart Association functional classification system<sup>13</sup> was used to rate the impact of symptoms of angina and dyspnoea on physical functioning. The classifications were as follows: 0=no symptoms, 1=symptoms only with strenuous effort, 2=symptoms on normal exertion, 3=symptoms on mild exertion, 4=symptoms at rest. The assessments were made by patients in response to a question delivered over the phone by the study coordinator.

### Psychological well being

The 30 item General Health Questionnaire (GHQ-30)<sup>14</sup> was used as a measure of patients' psychological well being.

### Life satisfaction

The Life Satisfaction Index A<sup>15</sup> was designed to measure five dimensions relating to enthusiasm for life, degree to which a person finds life meaningful and accepts personal responsibility for his life, congruence between desired and achieved goals, and positive mood. The primary rationale for using the Life Satisfaction Index A in this study was that it provided a measure of patients' attitude towards their life in general. This could ultimately be an important determinant of patients' self assessed utilities and consequently have implications for the re-calibration of the UBQ-H scoring system. The UBQ-H data should also be inversely correlated with life satisfaction scores—as health problems increase in severity, life satisfaction was expected to decline.

### RECRUITMENT AND PROCEDURES

Patients were contacted, informed of the study, and invited to participate—initially by a letter signed by their physician and then by a phone call from the study coordinator. Eligible patients had any symptoms of dyspnoea and angina rated at the time of this phone call. At baseline patients were sent a UBQ-H, the Life Satisfaction Index A, and some demographic questions and asked to return the form by reply paid envelope. About 10 days later patients were mailed a follow up UBQ-H, and the GHQ-30. A random 50% of the follow up UBQ-H questionnaires did not contain the time trade off question, and none contained the rating scale.

### Results

Of the 464 cardiovascular disease patients approached to participate in the study, 120 were excluded because their English was poor, 16 patients refused to participate, and six were excluded for other reasons (for example,

Table 1 UBQ items (baseline n=296)

UBQ-H items (response categories)	% Missing or N/A	% Indicating some impairment	Quartiles			Score for 90 <sup>th</sup> percentile
			LQ	M	UQ	
<i>Psychological distress (0-10)</i>						
Breathlessness	0	70	0	3	5	7
Difficulty sleeping	0	60	0	2	5	7
Lack of energy	0.3	72	0	3	5	7
Pain	0.3	52	0	1	3	5
Depression	0	55	0	1	4	6
Anxiety	0.3	61	0	2	5	6
Weight/appearance	0	53	0	1	5	7
Uncertainty	0.3	60	0	2	5	8
Anger	0	41	0	0	3	5
Loneliness	0	33	0	0	2	5
Loss of self confidence	0	41	0	0	3	5
Feeling dependent	0	38	0	0	2	6
Feeling sick	0	48	0	0	3	5
Muscle aches/pains	0	70	0	2	5	7
Difficulty concentrating	0	52	0	1	3	6
Other problems	4.7	46	0	0	3	6
<i>Self care (1-3)</i>						
Washing	0.3	10	1	1	1	2
Dressing	0.7	12	1	1	1	2
Eating	0.7	4	1	1	1	1
Using the toilet	0.7	4	1	1	1	1
<i>Social/usual activities (1-4)</i>						
Usual daily activities	0.3	40	1	1	2	2
Social life	0.3	37	1	1	2	2
Seeing friends/relatives	0.7	26	1	1	2	2
Leisure/hobby activities	0.7	39	1	1	2	3
Sex life	16.6	60	1	2	3	4
<i>Physical ability (1-4.5)</i>						
Mobility (1-6)	0.7	31	1	1	2	2
Climbing stairs (1-4)	0	51	1	1	2	3
Walking 500 metres (1-4)	0	47	1	1	2	3
Vigorous activities (1-4)	3.7	90	3	3	4	4
<i>Overall quality of life</i>						
Self rated health (1-4)	0	92	2	3	3	3
Time trade off (0.0-1.0)	9.2	58	0.67	0.96	1.00	0.53*
Rating scale (0-100)	4	97	50	70	85	35*
<i>Rosser</i>						
Disability (1-7)	0	52	1	2	3	4
Distress (1-4)	0	76	2	2	3	4
Utility (-1.486-1)	0	80	0.96	0.99	1.00	0.91*

\*A higher score on these items indicates better quality of life and so the lower 10<sup>th</sup> percentile score is shown.

severely impaired by stroke, Down's syndrome, etc). Three hundred and twenty two patients were consequently recruited to the study. The mean age of the sample was 60 years (SD=10), and 27% were women. Two hundred and sixty five were recruited from the general cardiac outpatient clinic, and 57 were recruited from the heart failure/transplant outpatient clinic. Thirty seven per cent reported no symptoms of angina or dyspnoea, 18% reported symptoms on strenuous exertion, 17% reported symptoms on normal exertion, 10% reported symptoms on mild exertion, and 18% reported symptoms even at rest.

Six principal investigations were performed. Firstly, descriptive statistics were calculated for questionnaire return rates, item completion rates, and item and scale scores. Secondly, the homogeneity of items within each domain was examined using Cronbach's  $\alpha$  and item-total correlation analyses. Thirdly, the validity of assigning an equal weight to items within a single quality of life domain was investigated using regression techniques. Fourthly, the reliability of the items and scales was measured using intraclass test-retest methods. Fifthly, the construct validity of the UBQ-H was investigated by testing for relations between it and other indicators of quality of life. Sixthly, the instrument's responsiveness to health changes resulting from various adverse events, such as myocardial infarction, was evaluated.

#### QUESTIONNAIRE RETURN AND ITEM COMPLETION RATES

Ninety two per cent of the baseline and 89% of the follow up UBQ-H questionnaires were mailed back (87% returned both questionnaires). The proportion of unusable responses for the individual baseline UBQ-H items was very low (column 2 of table 1). The items with the greatest proportion of unusable responses were: vigorous activities (4%, with 3% answering "not applicable"), rating scale (4%), other problems (5%), time trade off (9%), and sex life (17%, with 12% "not applicable"). Patients without a spouse/partner were more likely to have an unusable response to the sex life item ( $\chi_1^2=41.7$ ,  $p<0.0001$ ).

#### RANGE OF SCORES

At least 50% of patients reported some impairment on half of the items, but the percentiles shown in table 1 reveal that the data were generally skewed towards more favourable scores. Floor effects were pronounced for the self care items. Almost all patients in the sample had no difficulty with eating or using the toilet, but 10% reported at least some difficulty dressing and washing. Other items with considerably skewed distributions included: mobility, anger, loneliness, loss of self confidence, feeling sick, pain, usual daily activities, social life, seeing friend/relatives. Ninety per cent of all Rosser Utility scores fell between 0.91 and 1.00. The

Table 2 Test-retest reliability of UBQ-H

UBQ-H data (Cronbach's $\alpha$ )	Intra-class reliability			Follow up standard deviation	
	All patients (n=279*)	Stable health (n=217)	Unstable health (n=60)	Stable health (n=217)	Unstable health (n=60)
<i>Psychological distress</i> ( $\alpha=0.91$ )	0.80	0.81	0.79	1.7	2.1
Breathlessness	0.81	0.82	0.80		
Difficulty sleeping	0.81	0.81	0.82		
Lack of energy	0.73	0.73	0.71		
Pain	0.74	0.78	0.62		
Depression	0.67	0.69	0.60		
Anxiety	0.68	0.72	0.55		
Weight/appearance	0.83	0.83	0.83		
Uncertainty	0.72	0.73	0.69		
Anger	0.64	0.65	0.63		
Loneliness	0.72	0.80	0.35		
Loss of self confidence	0.69	0.72	0.62		
Feeling dependent	0.71	0.70	0.74		
Feeling sick	0.61	0.59	0.65		
Muscle aches/pains	0.75	0.76	0.74		
Concentration	0.76	0.72	0.88		
Other problems	0.54	0.61	0.37		
<i>Self care</i> ( $\alpha=0.79$ )	0.72	0.65	0.79	0.2	0.3
Washing	0.74	0.73	0.76		
Dressing	0.61	0.70	0.48		
Eating	0.62	0.51	0.79		
Using the toilet	0.74	0.65	0.85		
<i>Usual/social activities</i> ( $\alpha=0.85$ )	0.67	0.71	0.58	1.5	1.6
Usual daily activities	0.57	0.58	0.51		
Social life	0.57	0.61	0.46		
Seeing friends/relatives	0.56	0.59	0.47		
Leisure/hobby activities	0.62	0.64	0.55		
Sex life	0.79	0.84	0.58		
<i>Physical ability</i> ( $\alpha=0.80$ )	0.74	0.76	0.68	0.8	1.8
Mobility	0.63	0.61	0.67		
Climbing stairs	0.69	0.76	0.43		
Walking 500 metres	0.76	0.81	0.61		
Vigorous activities	0.53	0.51	0.61		
<i>Overall health items</i>					
Self rated health	0.74	0.75	0.73	0.7	0.9
Time trade off	0.66	0.69	0.62	0.2	0.2
<i>Rosser</i>					
Disability	0.66	0.67	0.64		
Distress	0.76	0.79	0.69		
Utility	0.61	0.56	0.72	0.04	0.07

\*=279 of 322 returned both UBQ-Hs, and two did not answer the item on health changes since baseline.

broadest range of responses were recorded for the items on: uncertainty, breathlessness, lack of energy, muscle aches/pains, difficulty sleeping, weight/appearance, sex life, and vigorous activities.

#### HOMOGENEITY OF SUMMARY SCALES

Cronbach's  $\alpha$  for the internal consistency of the domain specific summary scales ranged between 0.79–0.91 (see column 1 in table 2). The homogeneity of the scales was not appreciably increased with the removal of any item. An item-total analysis showed each item to be most highly correlated with its hypothesised quality of life domain. The average correlation between each item and its hypothesised scale was 0.62. The average correlation between each item and all other scales was 0.37. The psychological distress item on breathlessness was equally well correlated to the physical ability scale (0.60). Several of the items on social/usual activities and physical ability were

reasonably well correlated with either scale. The average correlation between these two types of items and their hypothesised scale was 0.68, and the correlation between the items and the alternative scale was 0.57.

#### IMPORTANCE OF ITEMS WITHIN DOMAINS

A common method for summarising information from items that fall within a particular quality of life domain is to sum or average the individual scores. This not only assumes that the items are homogenous measures of a single domain, but that each item is of equal subjective importance. This latter assumption was investigated by regressing the items from a domain against the rating scale measure of overall quality of life. The coefficients from the resultant model were used to estimate the subjective importance placed on each item within the domain. The correlation between the weighted and unweighted combinations of items for each domain was consistently very high (mean=0.95). The weighted combinations were moderately stronger predictors of overall quality of life, with  $r^2$  estimates about 4 to 9 percentage points greater. The relative reduction in variance explained by the unweighted averages compared with the weighted combinations ranged between 12% to 23%. The domain specific scale scores in subsequent analyses were calculated by taking an unweighted average of items.

Table 3 Validity of UBQ-H (Spearman correlations)

Baseline UBQ-H data	Physical		Psychological	
	Dyspnoea grade	Angina grade	Life satisfaction	GHQ-30
Physical ability	0.45	0.22	-0.28	0.43
Psychological distress	0.44	0.32	-0.53	0.69
Self care	0.25	0.12	-0.22	0.27
Social/usual activities	0.46	0.20	-0.30	0.55

All p values for correlations <0.005.

TEST-RETEST RELIABILITY

The median time between completion of the two UBQ-H assessments was 12 days. The test-retest reliability of the UBQ-H data was calculated for: (1) all patients; (2) for patients who reported on the follow up UBQ-H that their health was “about the same” as it was at baseline, and (3) for patients who did report a change in health. Table 2 shows the results. The median reliability of items and scales for patients who reported their health as remaining stable was 0.71 (lower quartile = 0.62, upper quartile = 0.76). The test-retest reliability of the Rosser

Utility score was lower than most items at 0.61, but the rank correlation between test and retest assessments for the Rosser Utility score was 0.83. Table 2 also lists the standard deviations for the principal indices from the UBQ-H at the follow up assessment. Scores from patients who reported a change in health tended to have a broader distribution than those from patients who reported their health had remained stable.

VALIDITY OF UBQ-H

The validity of a quality of life instrument cannot be unequivocally proved because there

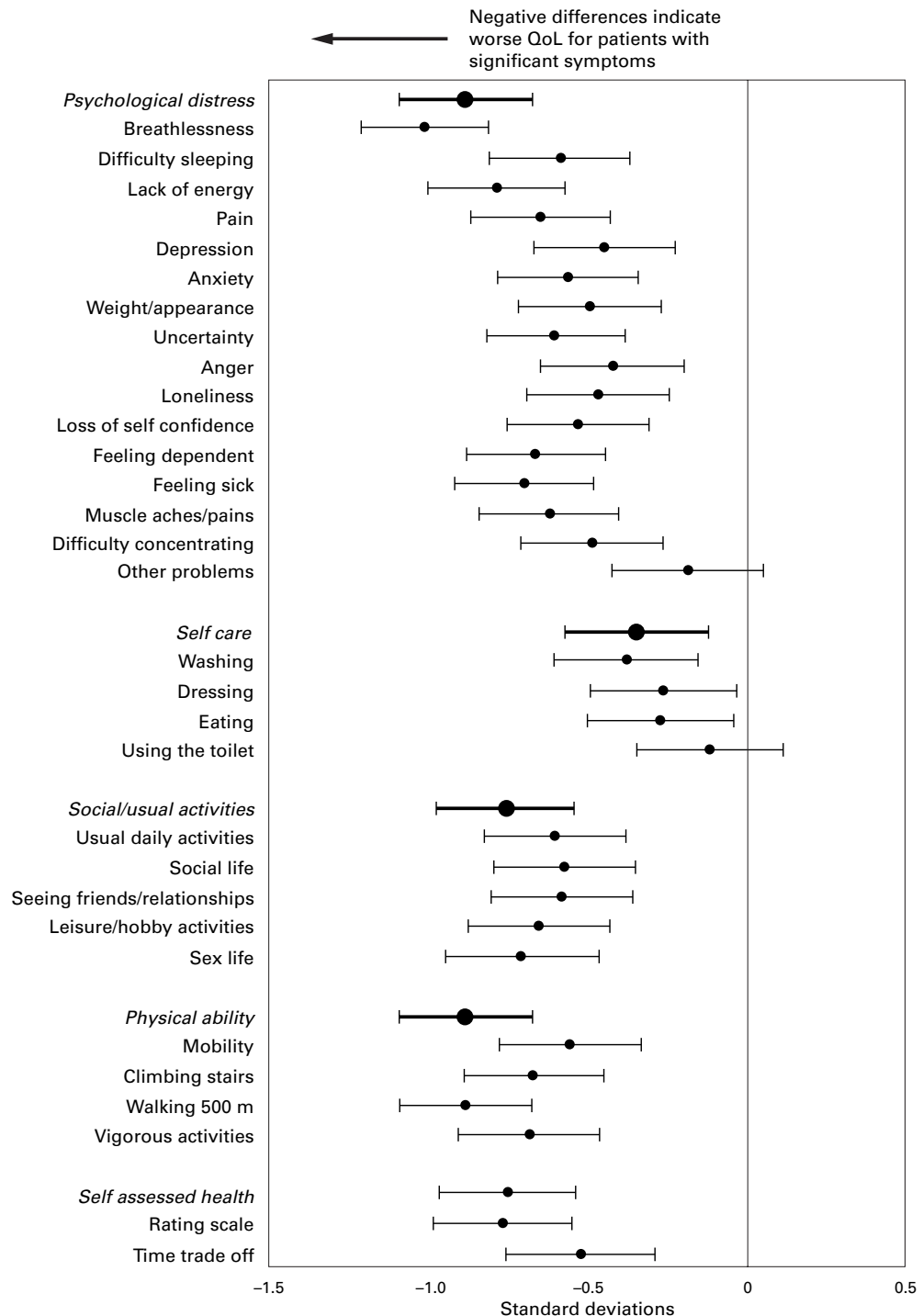


Figure 2 UBQ-H data for patients with/without significant symptoms.

exists no gold standard against which to judge it directly. Various indirect methods of testing an instrument's validity, such as construct validity, are therefore used. Construct validity involves testing for a relation between the instrument and another measure that is hypothesised to be related to the construct of interest. This can involve testing for significant correlations between the instrument and other relevant measures of the construct, or by contrasting groups that are expected to differ in terms of construct being assessed. It was hypothesised that (1) the UBQ-H measures of

physical/functional ability (physical ability, social/usual activities) would correlate with the two measures of physical impairment from symptoms of angina and dyspnoea, and (2) the GHQ-30 would correlate with the psychological distress scale of the UBQ-H.

The UBQ-H data were significantly correlated with the other information on health and quality of life in the expected direction (see table 3). The GHQ data correlated most highly with the psychological distress scale (0.69) as did the life satisfaction data (-0.53). The measure of physical impairment from

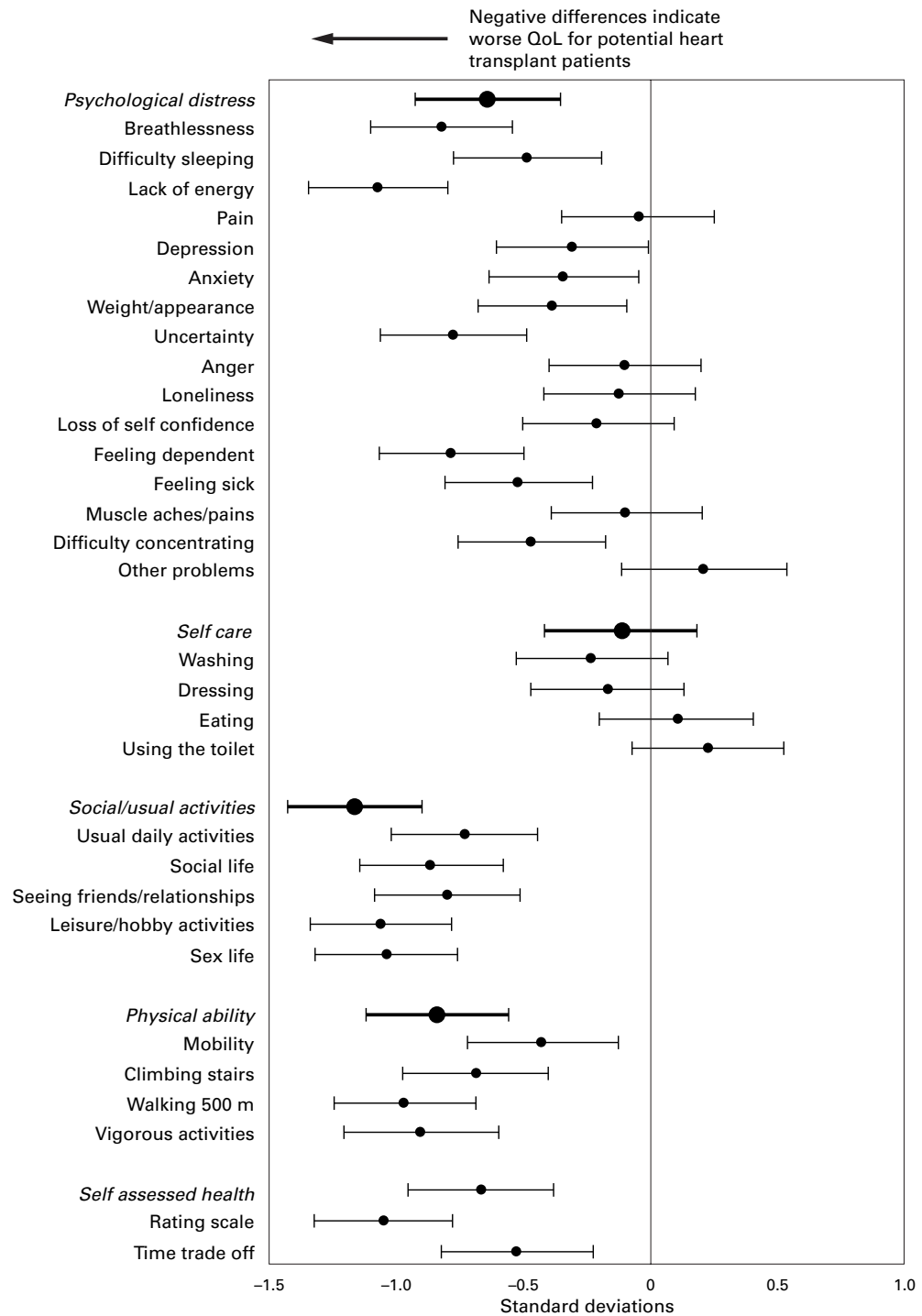


Figure 3 UBQ-H data for patients seeking/not seeking cardiac transplant.

Table 4 Mean difference between contrasted groups at year 3 on UBQ-H (p values calculated using Wilcoxon rank sum test)

Health status	UBQ-H domain specific summary scales				UBQ-H measures of overall quality of life			
	Physical ability	Self care	Social/usual activities	Psychological distress	Self assessed health	Rating scale	Time trade off	Rosser score
Myocardial infarction yes (n=38) no event (n=492)	0.31 p=0.002	0.01 p=0.8	0.31 p=0.001	0.93 p<0.0001	0.34 p=0.02	-7.26 p=0.003	-0.03 p=0.5	-0.05 p=0.001
Unstable angina yes (n=150) no event (n=492)	0.40 p<0.0001	0.02 p=0.3	0.28 p<0.0001	0.77 p<0.0001	0.37 p<0.0001	-8.96 p<0.0001	-0.04 p=0.2	-0.04 p<0.0001
Cancer yes (n=62) no event (n=492)	0.44 p<0.0001	0.06 p=0.3	0.38 p<0.0001	0.83 p<0.0001	0.54 p<0.0001	-9.72 p<0.0001	-0.06 p=0.01	-0.06 p<0.0001
Stroke yes (n=18) no event (n=492)	0.84 p<0.0001	0.27 p=0.01	0.69 p<0.0001	1.13 p=0.07	0.38 p=0.09	-15.37 p=0.004	-0.07 p=0.1	-0.11 p=0.02
Heart failure yes (n=34) no event (n=492)	0.53 p<0.0001	0.05 p=0.3	0.41 p<0.0001	0.77 p=0.02	0.51 p=0.002	-12.93 p<0.0001	-0.10 p=0.06	-0.07 p<0.0001
Hospitalisation yes (n=516) no event (n=492)	0.32 p<0.0001	0.04 p=0.08	0.27 p<0.0001	0.56 p<0.0001	0.29 p<0.0001	-7.79 p<0.0001	-0.05 p=0.002	-0.04 p<0.0001

A p value below 0.003 has a p value of <0.05 after adjusting for the 48 tests using the Hochberg<sup>23</sup> method.

symptoms of dyspnoea was about equally strongly correlated with the social/usual activities, physical ability, and psychological distress scales (around 0.44). The angina data were not highly correlated with any of the UBQ-H data. The self care scales did not correlate highly with any of the other measures of quality of life.

Two contrasted groups analyses were next performed. Firstly, UBQ-H data for patients with significant impairment from angina and/or dyspnoea ( $\geq$  grade 2) were compared with data from patients without significant impairment. Secondly, UBQ-H data from potential heart transplant candidates recruited from the transplant and heart failure clinic were compared with UBQ-H data from those patients recruited from the general cardiac outpatient clinic. It was hypothesised that: (1) patients with significant impairment from symptoms of angina and/or dyspnoea would score worse on the UBQ-H than other patients; and that (2) the potential heart transplant candidates would have greater health limitations and score worse on the UBQ-H than patients recruited from the general cardiac outpatient clinic.

Patients with significant symptoms of angina/dyspnoea (n=133) did score worse on all baseline UBQ-H items compared with other patients (n=163). A graph of the standardised difference between baseline UBQ-H scores for the two sub-groups illustrates this well (see fig 2). The difference between these two subgroups' scores was significant (95% confidence intervals did not cross 0) for all variables except for other problems, and using the toilet. The most significant differences between the two groups were detected by the items on breathlessness, walking 500 m and lack of energy, and the summary scales on psychological distress, physical ability and social/usual activities.

Figure 3 shows that patients from the transplant and heart failure clinic (n=54) scored significantly lower on all of the summary scales except self care than other patients (n=242). Variables that were best at discriminating between these two groups included the social/

usual activities scale, lack of energy, leisure/hobby activities, sex life, walking 500 m, vigorous activities, social life, breathlessness, and seeing friends. There were non-significant differences between the groups on anger, loneliness, other problems, lack of confidence, pain, muscle aches, and all self care items.

A second independent sample of 1112 patients taking part in the LIPID trial was also used to investigate the validity of the UBQ-H data. These patients received the UBQ-H as part of their regular assessment. At the time of this analysis, data had been collected at baseline, year one, and year three. Information on the occurrence of the following adverse events was available: myocardial infarction, unstable angina, stroke, cancer, hospitalisation, or cardiac failure. Table 4 compares domain specific UBQ-H scores at year 3 for LIPID patients who had experienced an adverse event since baseline versus those who had not. Patients who had experienced an event always obtained worse UBQ-H scores. These differences were almost always highly significant even after adjusting for multiple comparisons. The self care scale, and the time trade off were less effective in differentiating between the groups.

#### RESPONSIVENESS

UBQ-H scores for LIPID trial patients who experienced no hospitalisations remained stable between baseline and year three for all variables but physical ability, which improved slightly (p=0.01), and the time trade off, which also improved (p<0.0001). Anticipated reductions in UBQ-H scores were detected for patients who experienced an adverse event (see table 5). More than half of the effects measured were statistically significant at the conventional p value level of 0.05. This number was substantially reduced after adjustment for multiple comparisons.

The time between events and the UBQ-H assessment was 19 months on average. This will have given some patients sufficient time to recover from acute events such as myocardial infarction. To explore this hypothesis, we repeated the longitudinal analysis in the subset

Table 5 Mean change from baseline (*p* values calculated using Wilcoxon signed rank sum test)

Health status	UBQ-H domain specific summary scales				UBQ-H measures of overall quality of life		
	Physical ability	Self care	Social/usual activities	Psychological distress	Self assessed health	Time trade off	Rosser score
Myocardial infarction (n=38)	0.11 p=0.3	-0.01 p=0.6	0.15 p=0.09	0.20 p=0.7	0.03 p=0.9	-0.12 p=0.02	-0.02 p=0.2
QoL within 12 months of MI (n=29)*	0.28 p=0.005	-0.03 p=1	0.28 p=0.003	0.41 p=0.3	0.21 p=0.2	-0.09 p=0.04	-0.01 p=0.001
Unstable angina (n=150)	0.12 p=0.02	0.02 p=0.03	0.09 p=0.01	0.12 p=0.3	0.04 p=0.7	-0.06 p=0.002	-0.02 p=0.07
Cancer (n=62)	0.24 p=0.01	0.04 p=0.2	0.25 p=0.004	0.43 p=0.002	0.31 p=0.004	-0.11 p=0.001	-0.04 p=0.002
Stroke (n=18)	0.57 p=0.02	0.18 p=0.04	0.48 p=0.007	0.79 p=0.07	0.16 p=0.4	-0.07 p=0.4	-0.07 p=0.04
Heart failure (n=34)	0.17 p=0.3	0.05 p=0.2	0.23 p=0.004	0.26 p=0.5	0.06 p=0.8	-0.11 p=0.04	-0.05 p=0.03
Hospitalisation (n=516)	0.15 p<0.0001	0.03 p=0.001	0.12 p<0.0001	0.16 p=0.03	0.05 p=0.2	-0.08 p<0.0001	-0.02 p<0.0001

A *p* value <0.002 has a *p* value of <0.07 after adjusting for the 42 tests using the Hochberg<sup>23</sup> method. \*These tests were not included in the adjustment for multiple comparisons.

of myocardial infarction patients that experienced their event within the 12 months preceding the UBQ-H assessment. The reductions in quality of life for this subset were generally greater and the *p* values for change improved (see table 5).

### Discussion

Utility-based quality of life assessments are required for estimating quality adjusted survival as a measure of treatment effectiveness. Unfortunately there is a paucity of practical instruments available, and none are specifically designed for cardiovascular patients. The modifications made to the Health Measurement Questionnaire have resulted in an assessment designed for cardiovascular patients that performed well against tests of reliability and validity. This result was an important first step before attempting to re-calibrate the scoring mechanism based on the Rosser Index.

The UBQ-H return rates and item completion rates were impressively high and suggest that the questionnaire items appeared acceptable and intelligible to patients. Ninety per cent of all the questionnaires sent out in this study were returned and rates of unusable data were below 1% for 85% of items, and below 5% for 94% of items. Greater missing data rates have been reported for other quality of life assessments. The median missing data rate for SF-36 items, for instance, was reported as 4% in a study with a large sample (n=3445) of patients.<sup>16</sup> About 9% of the sample did not answer the time trade off item, however suggesting that the item may pose difficulties for some patients.

Most item scores were skewed towards values indicative of better quality of life. Yet, several items with even pronounced floor effects (for example, feeling dependent, and feeling sick) were still effective in discriminating between contrasted groups of patients. A future version of the UBQ-H could be pruned of items that were highly skewed and weakly correlated with overall quality of life. The four items on self care, for example, could be collapsed into a single item and included within an existing scale.

The items within domain specific scales were homogenous. Cronbach's *a* scores ranged

between 0.79–0.91 for the various scales and are similar to those obtained for the SF-36 scales with coronary artery disease patients (range: 0.74–0.92).<sup>16</sup> Each item was most strongly correlated with the domain that it was hypothesised to sample, compared with the alternative domains. Several of the social/usual activities and physical ability items were reasonably well correlated with either scale, and the psychological distress item on breathlessness was equally well correlated with the physical ability scale. Even if different scales sample unique quality of life domains, some inter-correlation must be expected because health impairments are liable to impact across multiple quality of life domains. A lower degree of intercorrelation between the social/usual activities and physical ability items would have been preferable to empirically support the hypothesised distinction between the domains. At present we recommend maintaining the distinction. This is useful for descriptive purposes and may be important in the development of an improved utility-based scoring system over the Rosser Index.

The intra-class test-retest reliability of the UBQ-H domain specific scales was consistent with other popular quality of life instruments. Reliability estimates ranging between 0.65 to 0.87 have been reported for the SF-36 scales,<sup>17</sup> and estimates between 0.75 and 0.88 have been reported for the Nottingham Health Profile.<sup>18</sup> The reliability of the Rosser scores was greater in the unstable health group compared with the stable health group. This paradoxical finding can be explained by the fact that between patient variance was greater for the unstable health group compared with the stable health group. Because reliability coefficients represent the ratio of the between patient variance to the total (error and between patient) variance, they will be larger if between patient variance is increased relative to error variance. The between group variance for the unstable health group was larger, and so the error component contributed less to total variance. This inflated reliability estimates somewhat. The variance was smaller in the stable health group because they were generally in better health with less potential for fluctuations in health.



The construct validity of the UBQ-H was supported by the significant correlations between it and other information on quality of life. The psychological distress scale correlated best with the Life Satisfaction data and GHQ data. The measure of physical impairment from symptoms of dyspnoea was equally well correlated to the physical ability, social/usual activities, and the psychological distress scales. The relation between the psychological distress scale and the symptom scales may be inflated because the psychological distress scale includes items of a somatic nature (for example, distress from pain) as well as items of an affective nature (for example, distress from feeling sad or depressed). The fact that the measure of physical impairment from symptoms of angina was not highly correlated with any of the UBQ-H data can be explained in part by the narrow distribution of angina scores. Seventy per cent of all patients had no limitation from symptoms of angina at all.

Patients with significant symptoms of angina and/or dyspnoea scored worse on all items and scales, especially on items that corresponded closely to the criteria used to differentiate the groups (for example, breathlessness). In the comparison between patients seeking/not seeking heart transplant, items reflecting impairments resulting from mechanical failure of the heart rather than coronary artery disease tended to be more important discriminators. For example, items on lack of energy, leisure/hobby activities, and sex life were more relevant than those on pain and muscle aches.

The items on physical ability and overall quality of life that were added to the Health Measurement Questionnaire to construct the UBQ-H were found to be significant correlates of health status. The effect sizes measured in the contrasted groups analyses using the items on climbing stairs, walking, and vigorous activities were in each case greater than that for the original mobility item. The effect sizes for the self assessed health, rating scale, and time trade off measures of overall quality of life were among the highest estimated. The additional item on muscle aches/pains was found to be useful in the comparison between patients with and without significant symptoms of angina or dyspnoea, but less so in the comparison between patients seeking/not seeking heart transplant.

The UBQ-H was able to differentiate between LIPID trial patients who had experienced an adverse event compared with those that had not, and was also responsive to changes in health within patients. The biggest effect estimates for change tended to be for events with longer term impact on quality of life such as stroke and cancer, compared with more acute events such as myocardial infarction. Several of the longitudinal analyses involved smaller numbers (for example, 34 heart failure) and the effects measured were not always statistically significant. Where numbers were larger, such as with the hospitalisation group, even smaller estimates of effect were highly significant. The length of time between the event and the administration of

#### KEY POINTS

- A cardiovascular extension of the Health Measurement Questionnaire was developed.
- The new instrument was found to provide reliable and valid data, and was responsive to changes in health states over time.
- The data collected by the instrument may be used to estimate a summary utility index for overall quality of life.
- The instrument will facilitate the calculation of quality adjusted survival time in treatment studies with cardiovascular patients.

the UBQ-H also played a part in the observed responsiveness of the instrument. The deleterious effect of myocardial infarction was more readily observable in the sub-group group of 29 patients who completed the year three UBQ-H assessment within 12 months of their event.

The regression analyses suggest that patients on average do not assign an equal importance to each item within the quality of life domains. The unweighted scales were not very different to those generated by the regression analysis and were used in this study in the interests of parsimony and the generalisability of results. Some method of weighting individual items that fall within a single scale could nevertheless be further pursued.

Given the UBQ-H's potential as a reliable and valid quality of life instrument, the development of an improved scoring system over the Rosser Index will now be undertaken. The nature of the relation between the UBQ-H scales and patients' self assessed utilities will be determined using regression techniques. The relation is unlikely to be linear additive. Other studies that have examined this type of question have found interactions between quality of life domains,<sup>19-21</sup> and the relation between value scales and utility scales is generally non-linear because of risk aversion.<sup>22</sup> Non-linear relations can be modelled using regression by transforming variables and/or fitting interaction parameters. The best predictive regression equation developed will be contrasted with the Rosser Index as alternative mechanism for translating UBQ-H data into utilities.

The UBQ-H will not necessarily be the optimal instrument for measuring the quality of life of cardiovascular patients in all settings. Once the re-scaling is finalised however, the UBQ-H should provide researchers with a practical method for obtaining utility estimates from cardiovascular patients. This is vital in cases where quality adjusted survival is a chosen measure of treatment effectiveness, and where a cost utility analysis is planned.

We would like to thank Royal Prince Alfred Hospital and St Vincent's Hospital for their assistance in recruiting patients.

Funding: funding for this study was provided by the National Heart Foundation, Bristol-Myers-Squibb, and the National Health and Medical Research Council.

Conflicts of interest: none.

- 1 Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989;5:559–75.
- 2 The LIPID Study Group. Design features and baseline characteristics of the LIPID (Long-Term Intervention with Pravastatin in Ischemic Disease) Study: a randomized trial in patients with previous acute myocardial infarction and/or unstable angina pectoris. *Am J Cardiol* 1995;76:474–9.
- 3 Kind P, Gudex C. Measuring health status in the community: a comparison of methods. *J Epidemiol Community Health* 1994;48:86–91.
- 4 Glasziou PP, Bromwich S, Simes RJ. Quality of life six months after myocardial infarction treated with thrombolytic therapy. *Med J Aust* 1994;161:532–6.
- 5 Rosser R, Kind P. A scale of valuation of states of illness: is there a social consensus? *Int J Epidemiol* 1978;7:347–58.
- 6 Kind P, Rosser R. The quantification of health. *European Journal of Social Psychology* 1988;18:63–77.
- 7 Magee TR, Scott DJ, Dunkley A, et al. Quality of life following surgery for abdominal aortic aneurysm. *Br J Surg* 1992;79:1014–16.
- 8 Rawles J, Light J, Watt M. Quality of life in the first 100 days after suspected acute myocardial infarction—a suitable trial endpoint? *J Epidemiol Community Health* 1992;46:612–16.
- 9 Norman-Taylor FH, Palmer CR, Villar RN. Quality-of-life improvement compared after hip and knee replacement. *J Bone Joint Surg Br* 1996;78:74–7.
- 10 Moinpour MC, Feigl P, Metch B, et al. Quality of life end points in cancer clinical trials: review and recommendations. *J Natl Cancer Inst* 1989;81:485–95.
- 11 EuroQoL Group. EuroQoL—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- 12 Ware JE, Sherbourne CD. The MOS 36-Item Short Form Health Survey (SF-36). I: conceptual framework and item selection. *Med Care* 1992;30:473–83.
- 13 New York Heart Association. Criteria Committee. *Nomenclature and criteria for diagnosis of diseases of the heart and blood vessels*. Boston: Little, Brown, 1979.
- 14 Goldberg D. *Manual of the General Health Questionnaire*. Windsor, England: NFER Publishing, 1978.
- 15 Neugarten B, Havighurst R, Tobin S. The measurement of life satisfaction. *J Gerontol* 1961;16:134–43.
- 16 McHorney CA, Ware JE, Lu JFR, et al. The MOS 36-item Short-Form health survey (SF-36): III. tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40–66.
- 17 Andresen EM, Patrick DL, Carter WB, et al. Comparing the performance of health status measures for healthy older adults. *J Am Geriatr Soc* 1995;43:1030–4.
- 18 Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract* 1985;35:185–8.
- 19 Torrance GW, Furlong W, Feeny D, et al. Multi-attribute preference functions: Health Utilities Index. *Pharmacoeconomics* 1995;7:503–20.
- 20 Sintonen H. *The 15-D measure of health-related quality of life. II feasibility, reliability and validity of its valuation system*. Working Paper 42 National Centre for Health Program Evaluation. Melbourne: 1995.
- 21 Veit CT, Rose BJ, Ware JE. Effects of physical and mental health on health-state preferences. *Med Care* 1982;20:386–401.
- 22 Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-Economic Planning Science* 1976;10:129–36.
- 23 Wright SP. Adjusted p-values for simultaneous inference. *Biometrics* 1992;48:1005–13.