

Refereed paper

Evaluation of a new short generic measure of health status: *howRu*

Tim Benson BSc BA MSc

Director, Abies Ltd; Interoperability Architect, Wellcome Trust Sintero Server Project, Cardiff University School of Computer Science and Informatics; and Honorary Senior Research Fellow, CHIME, University College London, London, UK

Steve Sizmur PhD

Survey Statistician, Picker Institute Europe, Oxford, UK

Justin Whatling MBBS BSc

Visiting Professor, CHIME, University College London, London, UK

Seref Arikan MSc

PhD Student, CHIME, University College London, London, UK

Dianna McDonald MA

Survey Team Manager, Picker Institute Europe, Oxford, UK

David Ingram PhD

Professor, CHIME, University College London, London, UK

ABSTRACT

Background Quality of life is paramount for patients and clinicians, but existing measures of health were not developed for routine use.

Objectives This paper describes the development and testing of a new generic tool for measuring health related quality of life (HRQoL) with direct comparison to the SF-12 Health Survey.

Methods The new tool (*howRu*) has four items (discomfort, distress, disability and dependence), rated using four levels (none, a little, quite a lot and extreme), providing 256 possible states (4^4); it has an aggregate scoring scheme with a range from 0 (worst) to 12 (best). Psychometric properties were examined in a telephone survey, which also recorded SF-12.

Results The *howRu* script is shorter than SF-12 (45 words vs 294 words) and has better readability statistics. 2751 subjects, all with long-term conditions (average age 62, female 62.8%), completed the survey; 21.7% were at the ceiling (no reported problems on any dimension); 0.9% at the floor. Inter-item correlations, Cronbach's alpha and principal factor analysis suggest that a single summary

score is appropriate. Correlations between the physical and mental components of both *howRu* and SF-12 were as expected. Across all patients the *howRu* score was correlated with PCS-12 ($r=0.74$), MCS-12 ($r=0.49$) and the sum of PCS-12 and MCS-12 ($r=0.81$). Subjects were classified by *howRu* score, primary condition, the number of conditions suffered, age group, duration of illness and area of residence. Across all six classifications, the correlation of the mean *howRu* score with the mean PCS-12 for each class was $r=0.91$, with MCS-12, $r=0.45$ and with the sum of PCS-12 and MCS-12, $r=0.97$.

Conclusions *howRu* is a new short generic measure of HRQoL, with good psychometric properties. It generates similar aggregate results to SF-12. It could provide a quick and easy way for practitioners to monitor the health of patients with long-term conditions.

Keywords: health status, health related quality of life, *howRu*, outcome assessment, patient reported outcome measures, SF-12

Introduction

Health cost inflation and ageing populations are driving health services to improve quality based on the results delivered to patients. A missing ingredient in the effort to span the quality chasm between what is done and what is possible¹ is our ability to measure the effectiveness of care, as perceived by the patient, using patient-reported outcome measures (PROMs).² The challenge is to measure health in a practical, generic way that is applicable to all conditions and care settings and delivers immediate feedback to patients, clinicians and managers.³

The benefits of routine measurement of HRQoL include helping to screen for problems, promoting patient-centric care, aiding patients and doctors to take decisions, improving communication amongst multidisciplinary teams, monitoring progress of individual or groups of patients and the quality of care in a population.^{4,5} However, in spite of demonstrated benefits, routine HRQoL assessment in day-to-day practice remains rare.⁶

The core concepts are not new. Encouraged by thinkers such as Donabedian,⁷ health status research began during the late 1960s with the application of operations research and systems thinking to health-care evaluation and resource allocation. One set of developers focused on the valuation of health states for policy and economic evaluation of healthcare programmes, but devoted little attention to the practicalities of data collection.⁸⁻¹⁰ Others developed lengthy health profiles to be completed by patients, leading to the term patient-reported outcome (PRO).^{11,12}

A second generation of generic instruments was introduced from the early 1990s, which included the Short Form SF-36¹³ and its derivatives, such as the SF-12 used in this study, the Euroqol EQ-5D¹⁴ and the Health Utilities Index (HUI).¹⁵ These methods were designed primarily for population surveys and clinical research.

We recognised the need for a new generation of instrument that would be shorter, quicker to use and designed for electronic data collection.

Method

Development

The concept to be measured is patients' perception of their HRQoL, defined as the impact of their health conditions and treatment on daily life.¹⁶ Conceptually, HRQoL assessments record patients' perceptions of their current health status in terms of how they feel

and how much they can do. Every HRQoL measurement instrument has two parts, a descriptive system (usually a questionnaire) and a scoring system.

Descriptive system

The name of the instrument, *howRu*, stems from the first question that a clinician may ask at a consultation, namely: 'How are you?'. Our approach is that of assessment at the time, rather than recall. Assessment captures the presence, absence, severity or intensity of a concept, while recall is based on recollection and memory, which is less reliable.¹⁷

The origins of the descriptive system can be traced to the work of Rachel Rosser, who more than 40 years ago developed a classification with eight classes of disability and four classes of distress, which she used to measure hospital output in terms of differences between admission, discharge and follow-up¹⁸ and to measure daily patient progress on hospital medical and surgical wards.¹⁹ Later, Rosser separated distress into separate axes for physical discomfort and emotional distress and disability into dysfunction and dependency.²⁰

The wording, design and scoring system of *howRu* evolved over a two-year period through numerous iterations, using pilot studies, feedback from colleagues and members of the public and desk research including literature review, dictionaries and thesauri. The purpose was to use simple terms and descriptions, in order to reduce the risk of ambiguity and to ensure that as many people as possible could use the measure reliably and consistently without training or support.

















The present descriptive system is illustrated in Figure 1 and has four items:

- 1 *Pain or discomfort* (short label: discomfort) is intended to cover the severity of physical symptoms including breathlessness, itching, dizziness and nausea
- 2 *Feeling low or worried* (distress) relates to emotional symptoms such as anxiety, stress, fatigue and depression
- 3 *Limited in what I can do* (disability) may include work, home and leisure activities (NB in an interview it is more appropriate to say 'you' rather than 'I')
- 4 *Dependent on others* (dependence) covers autonomy, self-care and other activities of daily living.

The severity of each item is rated using four levels (none, a little, quite a lot and extreme), which are indicated in mutually supporting ways to minimise cognitive load:

- Written labels: none, a little, quite a lot and extreme
- Colour: green, yellow, orange and red
- Position: increasing in severity from left to right
- Pictographs.

Circle one face on each line to tell us how you are today

How are you today?	None	A little	Quite a lot	Extreme
Pain or discomfort				
Feeling low or worried				
Limited in what I can do				
Dependent on others				

© 2008-2010 Abies Ltd. All rights reserved. 04/10 howRu™

Figure 1 *howRu* standard form 2010

The resulting matrix, with four items and four levels, provides 256 (4^4) different possible combinations of health state.

Scoring system

For analysis and reporting, each level is allocated a score on a 0–3 ordinal scale, with:

- Extreme = 0
- Quite a lot = 1
- A little = 2
- None = 3.

The overall *howRu* score is simply calculated by adding the scores for each item, giving a range from the floor, 0 ($4 \times$ extreme), to the ceiling, 12 ($4 \times$ none).

Healthcare IT systems use codes to identify measures used. Codes are required to enable interoperability between healthcare IT systems and meaningful use of electronic health record (EHR) systems; *howRu* is one of the first generic HRQoL instruments to have been allocated codes in LOINC,²¹ Read Codes Version 2,²² Clinical Terms Version 3²³ and SNOMED CT (UK extension)²⁴ (see Table 1).

Table 1 LOINC, Read V2, CTV3 and SNOMED CT (UK extension) codes for *howRu*

Rubric	LOINC	Read V2	CTV3	SNOMED CT UK extension (hierarchy)
<i>howRu</i> rating scale	55744–7	38DY.	XaQtg	515381000000104 (assessment scale)
<i>howRu</i> rating score	55749–6	16ZB.	XaQuy	515461000000100 (observable entity)
Pain or discomfort	55745–4	16ZB0	XaQuz	515481000000109 (observable entity)
Feeling low or worried	55746–2	16ZB1	XaQu0	515501000000100 (observable entity)
Limited in what I can do	55747–0	16ZB2	XaQu1	515521000000109 (observable entity)
Dependent on others	55748–8	16ZB3	XaQu2	515541000000102 (observable entity)

SF-12 Health Survey

The SF-12 Health Survey (Version 1.0)²⁵ which is used in this study was developed as an even shorter (12-item) version of the Short Form SF-36 (36 items) of RAND's Medical Outcomes Study core survey instrument (116 items). SF-12 has 12 items and the results are reported as two summary scores: the Physical Components Summary (PCS-12) and the Mental Components Summary (MCS-12). PCS-12 and MCS-12 are scaled so that the normal distribution of the US population has a mean of 50 and standard deviation of 10, so a score of 40 is one standard deviation below the norm.

Data collection

A validation study was undertaken to examine the psychometric properties and construct validity of *howRu* and to compare these with SF-12. The data was collected during the summer of 2008 as part of two telephone surveys (using the same questionnaire) into patients' experience of services for long-term conditions in two regions of the UK. In one survey, a sample of 1001 cases was drawn from one English county; in the second survey, the sample was 1907 cases from five counties. These surveys covered a wide range of socio-economic deprivation. The fieldwork used a standard script and computer-assisted telephone interviewing by a professional telephone survey company on behalf of the Picker Institute Europe. In each survey the sample was drawn using a random digit dialling procedure. Initial screening established whether the household contained an adult over the age of 16 with one or more long-term conditions from a list of 21 (angina, heart failure, high blood pressure, other heart condition, asthma, emphysema, bronchitis, other respiratory illness, depression, anxiety, other mental illness, arthritis, back pain, epilepsy, diabetes, cancer, recovering from a stroke, recovering from a heart attack and other chronic conditions).

Where two or more adults in the same household qualified, the interviewer asked to speak to a male (or the youngest person), as people from these groups are generally more difficult to obtain. Each study also gathered data on each subject's year of birth (used to calculate age), sex, locality, the presence or absence of each long-term condition, from which the number of conditions was calculated, the name and year of diagnosis of their primary condition (used to calculate duration of illness). The SF-12 items were asked before the *howRu* items. The telephone script version of *howRu*, used in the study is shown in Box 1.

Box 1 *howRu* telephone script used in survey

We'd now like to ask about how you are feeling TODAY and how much you can do.
How are you today?
– Do you have any of the following:
Symptoms, such as pain?
None; slight; quite a lot; extreme
Feeling low or worried?
None; slight; quite a lot; extreme
Limited in what you can do?
None; slight; quite a lot; extreme
Dependent on others?
None; slight; quite a lot; extreme

Each survey also contained additional questions, which are not reported here, and was about 2600 words in total. Subjects were informed that the questionnaire would take ten to 15 minutes to complete; information is not available about how long it took to complete each part of the questionnaire.

The data supplied were anonymous. The results from the two surveys were pooled for analysis. Ethical approval was not required because the surveys were anonymous surveys of the general public.

Validation

Construct validation is an investigation of score meaning, providing information about how scores may be interpreted and used.²⁶ We set out to test the following hypotheses.

- Correlations between the four *howRu* items would be moderate, averaging approximately between 0.4 and 0.5. These would be strongest amongst the three 'physical' items of *howRu* (discomfort, disability and dependence).
- The correlation between the three 'physical' items of *howRu* and the SF-12 Physical Components Summary, PCS-12, would be stronger than their correlation with the SF-12 Mental Components Summary, MCS-12.
- The correlation of the 'mental' item of the *howRu* (distress) would be stronger with MCS-12 than with PCS-12.
- The correlation of the *howRu* score, PCS-12 and MCS-12 with independent variables such as age, number of conditions and duration of illness would be similar.
- *howRu* and SF-12 (PCS-12 and MCS-12) would discriminate to a similar degree between patients

with different primary conditions and area of residence.

Analysis was performed using SPSS version 16.0.

Results

Length and readability

We used Microsoft Word 2008 readability statistics to measure the number of words, the readability grade and the reading ease scores for the *howRu* standard version (Figure 1), the *howRu* telephone script (Box 1) and the SF-12 script used in this study (see Table 2).

Missing values

The overall sample size was 2908. After excluding all subjects with any missing values, the sample was reduced to 2751. For SF-12 the missing value rate was 3.0%, meaning that 3% of respondents did not provide an answer to one or more of the SF-12 questions, which prevented calculation of the summary scores, PCS-12 and/or MCS-12. A further 2.4% of the sample had a missing value for one or more variables used in the analysis and these were also excluded. The missing value rate for *howRu* was zero (0%); that is, all respondents answered all four *howRu* questions.

Descriptive statistics

The average age was 61.9 years (SD 14.6 years), 63% women, 37% men. All respondents had at least one long-term condition; 59.5% reported having more than one condition (see Table 3). Conditions are listed in the order asked.

The ceiling state (*none* reported on all four items) accounted for 608 ratings (22.1%). The principle conditions of subjects who most frequently reported

at the ceiling were: high blood pressure (48.3% of subjects with high blood pressure as their principal condition reported no problems), high cholesterol (45.1%), asthma (35.3%) and diabetes (34.7%). On the other hand, less than 4% of patients with heart failure, varicose veins, emphysema, arthritis, back pain or recovering from a stroke were at the ceiling.

Twenty-five subjects (0.9%) reported being in the worst (floor) state (*extreme* reported on all four items). The most commonly reported primary conditions for these subjects were emphysema (5.7% of those with emphysema), back pain (4.3%) and arthritis (1.4%).

The overall frequency distribution for each of the *howRu* states is shown in Table 4. The range of response rates for each of the 16 cells was from 4.9% (*extreme distress*) to 58.4% (*no dependence*).

The ten most common states are shown in Table 5; these accounted for 47.2% of ratings. In all, 203 out of 256 possible states (79.3%) were used.

Internal structure

The internal structure of *howRu* was explored by examining the correlations between each pair of items (Table 6). All correlations were significant at the $P < 0.01$ level (2-tailed).

Principal factor analysis of the *howRu* items, using maximum likelihood extraction, generated factor loadings shown in Table 7 (mean 0.70). The loadings are the correlation between the extracted factor and each item. Cronbach's alpha was 0.80.

Comparison with SF-12

Table 8 shows the frequency distribution of subjects with each *howRu* score value, and the mean and standard deviation of the physical (PCS-12) and mental (MCS-12) components, summary scores and the sum of PCS-12 and MCS-12 (PCS-12 + MCS-12). The correlation of the *howRu* scores with the mean values of the PCS-12 score ($r=0.958$), MCS-12 ($r=0.986$) and PCS-12 + MCS-12 ($r=0.993$) are very high.

Table 2 Readability data

Instrument	Number of words	Flesch-Kincaid Grade Level	Flesch Reading Ease
<i>howRu</i> form	40	1.9	89
<i>howRu</i> telephone script	45	2.4	89
SF-12 script	294	8.4	68

Table 3 Overall distribution of subjects and those at ceiling (no problems on any dimension) and at floor (extreme on all dimensions)

Characteristic	No of subjects (%)	Subjects at ceiling (%)	Subjects at floor (%)
Gender			
Male	1023 (37.2)	242 (23.7)	5 (0.5)
Female	1728 (62.8)	354 (20.5)	20 (1.2)
Age range			
Under 30	84 (3.1)	23 (27.4)	0 (0)
30–39	126 (4.6)	43 (34.1)	3 (2.4)
40–49	300 (10.9)	82 (27.3)	5 (1.7)
50–59	515 (18.7)	101 (19.6)	3 (0.6)
60–69	809 (29.4)	181 (22.4)	6 (0.7)
70–79	678 (24.6)	135 (19.9)	6 (0.9)
80+	239 (8.7)	31 (13.0)	2 (0.8)
No. of conditions			
1	1113 (40.5)	353 (31.7)	9 (0.8)
2	748 (27.2)	159 (21.3)	7 (0.9)
3	440 (16.0)	57 (13.0)	1 (0.2)
4	221 (8.0)	19 (8.6)	3 (1.4)
>5	229 (8.3)	8 (3.5)	5 (2.2)
Duration of principal condition			
Less than 3 years	635 (23.1)	123 (19.4)	7 (1.1)
3–9 years	890 (32.4)	237 (26.6)	6 (0.7)
10–19 years	657 (23.9)	132 (20.1)	7 (1.1)
20+ years	569 (20.7)	104 (18.3)	5 (0.9)
Principal condition			
Angina	52 (1.9)	5 (9.6)	0 (0)
Heart failure	22 (0.8)	0 (0)	0 (0)
High blood pressure	443 (16.1)	214 (48.3)	0 (0)
High cholesterol	71 (2.6)	32 (45.0)	0 (0)
Other heart condition	81 (2.9)	20 (24.7)	0 (0)
Varicose veins	7 (0.3)	0 (0)	0 (0)
Asthma	232 (8.4)	82 (35.3)	0 (0)
Emphysema	35 (1.3)	1 (2.8)	2 (5.7)
Bronchitis	11 (0.4)	1 (9.1)	0 (0)
Other respiratory illness	40 (1.5)	2 (5.0)	1 (2.5)
Depression	55 (2)	7 (12.7)	0 (0)
Anxiety	19 (0.7)	2 (10.5)	0 (0)
Other mental illness	16 (0.6)	3 (18.8)	0 (0)
Arthritis	590 (21.5)	17 (2.9)	8 (1.4)
Back pain	139 (5.1)	5 (3.6)	6 (4.6)
Epilepsy	17 (0.6)	4 (23.5)	0 (0)
Diabetes	297 (10.8)	103 (34.7)	1 (0.3)
Cancer	70 (2.5)	12 (17.1)	1 (1.4)
Recovering from a stroke	32 (1.2)	1 (3.1)	0 (0)
Recovering from a heart attack	31 (1.1)	4 (12.9)	0 (0)
Other chronic condition	489 (17.8)	81 (16.6)	6 (1.2)
All conditions	2751 (100)	608 (22.1)	25 (0.9)

Table 4 Frequency distribution for *howRu* items

Level	None	Slight	Quite a lot	Extreme
Item				
Discomfort (symptoms such as pain)	1199 (43.6%)	626 (22.8%)	701 (25.5%)	225 (8.2%)
Distress (feeling low or worried)	1521 (55.3%)	663 (24.1%)	431 (15.7%)	136 (4.9%)
Disability (limited in what you can do)	1029 (37.4%)	695 (24.8%)	713 (25.9%)	314 (11.4%)
Dependence (dependent on others)	1608 (58.4%)	443 (16.1%)	500 (18.2%)	200 (7.3%)
All items	48.5%	21.8%	21.7%	8.1%

Table 5 Most commonly reported *howRu* health states

Discomfort	Distress	Disability	Dependence	<i>n</i>	%
None	None	None	None	596	21.7
Slight	None	None	None	130	4.7
None	None	Slight	None	125	4.5
Slight	None	Slight	None	100	3.6
None	Slight	None	None	100	3.6
Quite a lot	Quite a lot	Quite a lot	Quite a lot	63	2.3
Slight	Slight	Slight	None	51	1.9
None	Slight	Slight	None	49	1.8
Quite a lot	Slight	Quite a lot	Quite a lot	42	1.5
Quite a lot	None	Quite a lot	Quite a lot	42	1.5

Table 6 Pearson correlations between *howRu* items

	Discomfort	Distress	Disability	Dependence
Discomfort	–	0.40	0.58	0.47
Distress		–	0.45	0.39
Disability			–	0.65
Dependence				–

The correlation matrix between *howRu* items and SF-12 PCS-12, MCS-12 and PCS-12 + MCS-12 is shown in Table 9. As expected, the physical *howRu* items (discomfort, disability and dependence) were more highly correlated with the physical PCS-12 (mean $r=0.66$), than with the mental MCS-12 (mean $r=0.32$). The mental *howRu* item, distress, was more highly correlated with the mental MCS-12 ($r=0.59$)

than with the physical PCS-12 ($r=0.33$). The individual *howRu* scores for each subjects were correlated with PCS-12 ($r=0.74$), with MCS-12 ($r=0.49$) and with the sum of PCS-12 and MCS-12 ($r=0.81$).

Exploratory principal factor analysis on the *howRu* items, PCS-12 and MCS-12 (Table 10) suggested that discomfort, disability and dependence together with PCS-12 loaded substantially onto one factor, and that

Table 7 Primary factor loadings of each *howRu* item

Item	Loading
Discomfort	0.67
Distress	0.53
Disability	0.87
Dependence	0.74

distress and MCS-12 loaded onto the second factor. The loadings are the correlation between the factor and each item. The two factors correlated moderately with each other ($r=0.31$). These findings accord with the hypothesised relationships between these items.

Primary conditions

Table 11 shows the mean *howRu*, PCS-12, MCS-12 and PCS-12 + MCS-12 scores for each primary condition, sorted by *howRu* score rank. The correlations

Table 8 Frequency of each *howRu* score and mean SF-12 Physical Components Summary (PCS-12), Mental Components Summary (MCS-12), MCS-12 + PCS-12 and their standard deviations (SD)

<i>howRu</i> Score	No. of subjects	Cumulative %	Mean PCS-12 (SD)	Mean MCS-12 (SD)	Mean MCS-12 + PCS-12 (SD)
12	596	100.0	52.7 (5.9)	54.8 (6.7)	107.4 (8.4)
11	382	78.3	50.0 (7.4)	52.9 (8.0)	103.0 (9.3)
10	289	64.4	45.7 (9.6)	51.4 (9.8)	97.1 (11.5)
9	234	53.9	40.3 (10.1)	51.3 (10.2)	91.6 (11.6)
8	276	45.4	37.2 (10.2)	48.8 (11.5)	86.0 (12.5)
7	210	35.4	34.4 (10.3)	47.5 (11.6)	81.9 (12.4)
6	207	27.8	30.5 (8.5)	45.5 (12.2)	76.0 (12.5)
5	197	20.2	29.2 (9.1)	42.6 (12.7)	71.8 (12.4)
4	145	13.1	27.1 (8.1)	39.4 (12.8)	66.4 (10.6)
3	92	7.8	26.4 (7.8)	39.5 (13.3)	66.0 (12.1)
2	61	4.5	24.5 (6.4)	35.3 (11.6)	59.8 (9.5)
1	37	2.3	25.5 (7.7)	31.1 (9.6)	56.7 (11.9)
0	25	0.9	24.0 (3.3)	31.1 (8.1)	55.0 (6.9)
All scores	2751	100	40.7 (12.9)	48.9 (11.7)	89.6 (18.8)

The mean *howRu* score was 8.46 (SD 3.12)

Table 9 Correlations between *howRu* items and PCS-12, MCS-12 and PCS-12 + MCS-12

<i>howRu</i> item	PCS-12	MCS-12	PCS-12 + MCS-12
Discomfort	0.64	0.28	0.61
Distress	0.33	0.59	0.59
Disability	0.71	0.36	0.71
Dependence	0.64	0.35	0.65
<i>howRu</i> score	0.74	0.49	0.81

Table 10 Structure matrix of loadings for principal factor analysis

Item	Factor 1	Factor 2
Discomfort	0.69	0.34
Distress	0.42	0.71
Disability	0.82	0.43
Dependence	0.71	0.39
PCS-12	0.91	0.29
MCS-12	0.26	0.83

Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization, Kappa = 2

between the mean *howRu* score and mean PCS-12 ($r=0.70$) and mean MCS-12 ($r=0.52$) are high, while the correlation with the mean of the sum of PCS-12 and MCS-12 is very high ($r=0.97$). Pair-wise comparisons of the *howRu* scores between each pair of the 21 conditions (190 pairs) found that only one pair (diabetes and asthma) had a similar distribution that was likely to have occurred by chance (2-tail *t*-test, $P > 0.05$).

Other variables

Correlations between the mean values of the *howRu* score and the mean values of PCS-12, MCS-12 and PCS-12 + MCS-12 are shown in Table 12 for subjects grouped by number of conditions, age group, duration of illness and locality.

Table 11 Mean *howRu*, PCS-12 and MCS-12 scores for primary conditions ranked by mean *howRu* score

Primary condition	<i>n</i>	Mean <i>howRu</i> Score	Mean PCS-12	Mean MCS-12	Mean PCS-12 + MCS-12
High cholesterol	71	10.73	50.6	54.0	104.7
High blood pressure	443	10.71	50.3	52.6	102.9
Asthma	232	9.96	45.5	51.5	97.0
Diabetes	297	9.81	44.9	50.9	95.9
Varicose veins	7	9.57	39.4	56.4	95.8
Recovering from a heart attack	31	9.29	42.6	49.6	92.2
Other heart condition	81	8.99	40.6	52.0	92.6
Angina	52	8.48	38.7	48.5	87.2
Cancer	70	8.44	39.7	47.9	87.6
Heart failure	22	8.27	32.4	54.5	86.9
Epilepsy	17	8.18	46.8	39.8	86.6
Anxiety	19	7.95	52.6	32.8	85.4
Other chronic condition	489	7.66	38.4	46.8	85.2
Bronchitis	11	7.36	35.5	51.2	86.7
Depression	55	7.31	45.9	30.7	76.6
Other mental illness	16	7.06	41.8	35.2	77.0
Arthritis	590	6.87	33.7	48.7	82.4
Other respiratory illness	40	6.83	31.4	46.5	77.9
Emphysema	35	6.29	28.7	45.9	74.6
Back pain	139	6.17	32.6	45.5	78.1
Recovering from a stroke	32	6.03	32.4	43.2	75.6

Table 12 Correlation of the mean *howRu* score with PCS-12, MCS-12 and PCS-12 + MCS-12 for different subject categories

Category	No. of classes	Mean <i>howRu</i> score vs PCS-12 (<i>r</i>)	Mean <i>howRu</i> score vs MCS-12 (<i>r</i>)	Mean <i>howRu</i> score vs PCS-12 + MCS-12 (<i>r</i>)
<i>howRu</i> score	13	0.958	0.986	0.994
Principle condition	21	0.698	0.526	0.967
Number of conditions	5	0.983	0.967	0.998
Age group	7	0.905	-0.308	0.941
Duration of illness	4	0.996	-0.088	0.981
Area of residence	15	0.916	0.623	0.915
Mean correlation (<i>r</i>)	-	0.909	0.451	0.966

Discussion

Principal findings

This is the first published account of *howRu*, a new tool for measuring patient-reported health status. *howRu* measures health in terms of how the patient is feeling, physically and mentally, and how much they can do, in terms of loss of function and independence.

Evidence about the internal structure of the instrument suggests that it is appropriate to represent the results of the *howRu* questionnaire with a single score value. Inter-item correlations should ideally average between 0.4 and 0.5 for a relatively specific construct;²⁷ they should not be too high, which would imply that the items are asking the same question using different words, or too low, which would imply that items are about unrelated domains. The average inter-item correlation of the four *howRu* items was 0.50 (range 0.39 to 0.65), at the top end of the expected range. Principal factor analysis (average loading 0.70) and Cronbach's alpha (0.80) also suggest that the *howRu* items measure different aspects of an underlying continuum. Cronbach's alpha measures the extent to which items are consistent with each other and may be used together reliably as a single score. Alpha increases with the number of items in a scale (*howRu* has only four items) and should be in the range 0.70 and 0.90, depending on the measurement purpose.²⁸ An alpha of 0.80 is in the centre of the desired range for a scale of this length and is acceptable for individual-level measurement.^{29,30}

The scoring system used in *howRu* is a simple aggregate score in which a higher score indicates better health. This scoring system is similar to that used by the Apgar,³¹ Glasgow Coma³² and Oxford Hip and

Knee³³ scores. It is easy to use and understand and is transparent.

At the aggregate level, subjects were classified according to *howRu* score, primary condition, number of conditions, age group, duration of illness and area of residence. The correlation of the mean *howRu* score and the mean value of PCS-12 + MCS-12 across these six variables was very high, mean $r=0.966$ (range: 0.915–0.998), which provides strong support for the notion that *howRu* and SF-12 are measuring the same thing. The correlation of the mean *howRu* score and the mean value of physical components PCS-12 was also high, mean $r=0.909$ (range 0.698–0.996), but lower with mental components MCS-12, mean $r=0.451$ (range: 0.308–0.967).

Implications for practice

Barriers to use of any HRQoL measure include the respondent burden (the time needed to complete the forms) and the need for staff to be trained to understand the results.³⁴ An ideal system of health assessment needs to be clinically useful and timely, sensitive to change, culturally sensitive, low burden, low cost, involve the patient and built into standard procedures and needs to meet the requirements of regulators, payers and continuous quality improvement.³⁵ Changes in process, work flow and information systems are likely to be needed to ensure that the assessments are done regularly and to inform clinical decisions;³⁶ this requires technical support to set up and maintain the system (paper or electronic), help and explanations for individual patients, staff education and senior management backing.³⁷

Short questionnaires (parsimony) tend to provide higher participation rates, reduced respondent resistance

and better quality data, with less missing data and fewer satisficing responses (where a respondent answers questions in an unthinking way).³⁸ Readability is self-evidently important, yet Paz and colleagues³⁹ found that all commonly used instruments have many items with readability levels below the threshold recommended for documents to be used by vulnerable people of a readability grade score of 5.0 or less and a reading ease score of over 80.⁴⁰ The *howRu* form has a grade score of 1.8 and an ease score of 89. The acceptability of *howRu* is indicated by every respondent completing all parts of *howRu* (100% completion) in a telephone survey of almost 3000 subjects with long-term conditions.

howRu was designed to take no more than a few seconds using electronic data collection and integration with electronic patient records as part of other routine tasks that patients have to do, such as booking appointments, checking in on arrival at clinic, or ordering or collecting repeat medication. The variety of healthcare settings and processes requires a choice of paper, verbal and electronic modes of assessment. The user interface is suitable for touch screen consoles and portable devices such as the iPhone. Coding using industry standard coding schemes such as LOINC, Read Codes and SNOMED CT is essential for systems integration and data sharing; *howRu* is the first generic instrument to have adopted these standard clinical coding schemes, which facilitate feedback at the point of care and linkage to case mix and demographic data for longitudinal analysis.

Comparison with literature

There is no gold standard for generic HRQoL measures, but based on PubMed searches, the three most widely used instruments are SF-36 (8405 citations), SF-12 (1115 citations) and EQ-5D (1294 citations). These instruments and others have been reviewed in detail by McDowell⁴¹ and in reports for the UK Department of Health for the general population,⁴² long-term conditions⁴³ and elective surgery.⁴⁴

The overall correlations obtained in this study of the *howRu* score with the PCS-12 ($r=0.74$) and the MCS-12 ($r=0.49$) compare favourably to correlations of EQ-5D with PCS-12 (0.66) and MCS-12 (0.41) previously reported by Johnson and Pickard⁴⁵ in a general population survey in Canada. The correlation of the *howRu* score with the sum of PCS-12 and MCS-12 ($r=0.81$) is higher than any of the correlations reported by Hawthorne and colleagues between five generic utility instruments (AQoL, EQ-5D, HUI3, 15D and SF6D), which were in the range 0.66 to 0.80.⁴⁶

Limitations of method

The validation study was piggy-backed onto two studies to assess the health states of patients with long-term conditions which had previously been organised. The study population was limited to patients with long-term conditions, living in their own homes. These results are therefore not directly comparable to the general population or patients in institutions. Further work is required to test the instrument with different population groups, to develop population norms for the general population and to test the instrument in institutions and with older and young people.

The survey used a telephone interview, which may not be directly comparable with electronic or paper-based surveys. However, *howRu* and SF-12 data were collected in the same interview using the same method and so the comparisons between them are likely to be valid. It is possible that coloured pictographs, used on the paper and screen versions, could change responses, although other studies have shown that pictures made no difference to the results.⁴⁷

The standard version of *howRu* is shown in Figure 1, while the telephone survey used a telephone script based on an earlier beta version, dated July 2008 (Box 1). The main difference between the standard version and that used in this study is that the phrase *Pain or discomfort* has since replaced *Symptoms such as pain*.

The nature of the survey meant that we were not able to measure test-retest reliability, responsiveness and sensitivity to change, or to evaluate the instrument in different clinical settings. Further work is required to investigate these and other aspects of the instrument.

Conclusions

howRu is a new short generic measure of HRQoL, designed for routine clinical use at the point of care to provide immediate feedback to patients, clinicians and managers as part of continuous quality improvement, integrated with IT systems. The psychometric evidence from a telephone survey of patients with long-term conditions provides strong support for the validity of *howRu*. Although they are very different in design and construction, *howRu* and SF-12 give very similar results at the aggregate level.

CONTRIBUTIONS

TB, JW and SA developed the *howRu* instrument. TB wrote the paper with contributions from the other authors. SS was responsible for the psychometric

analysis and DM for data collection; DI supervised the MSc project of Timea Helter, who piloted the first prototype of *howRu*.

ACKNOWLEDGEMENTS

We wish to acknowledge the contributions of Amir Hannan, Andrew Lloyd, Angela Coulter, Brian Fisher, Caroline Powell, David Stables, Ed Conley, Ewan Davis, Henry Potts, James Case, Jeannette Murphy, Joanne Mak, Mary Hawking, Susan Horsewood-Lee and Timea Helter in the development of *howRu* and in preparing this paper, as well as the suggestions of anonymous reviewers of previous versions.

REFERENCES

- Institute of Medicine. *Crossing the Quality Chasm: a new health system for the twenty-first century*. Washington, DC: National Academy Press, 2001.
- Darzi. *High Quality Care for All: NHS next stage review final report*. London: Department of Health, 2008.
- NHS Outcomes, Performance and Productivity: Report of the Office of Health Economics Commission*. London: OHE, 2008.
- Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Quality of Life Research* 2009;18:115–23.
- Frost MH, Bonomi AE, Cappelleri JC et al. Applying quality-of-life data formally and systematically into clinical practice. *Mayo Clinic Proceedings* 2007;82:1214–28.
- Fung CH and Hays RD. Prospects and challenges in using patient-reported outcomes in clinical practice. *Quality of Life Research* 2008;17:1297–302.
- Donabedian A. Evaluating the quality of medical care. *The Millbank Memorial Fund Quarterly* 1966;44:166–203.
- Fanshel S and Bush JW. A health status index and its application to health-services outcomes. *Operations Research* 1970;18:1021–66.
- Torrance GW, Thomas W and Sackett D. A utility maximisation model for evaluation of health care programs. *Health Services Research* 1972;7:118–33.
- Culyer AJ, Lavers RJ and Williams A. Social indicators: health. *Social Trends* 1971;2:31–42.
- Ware JE. Scales for measuring general health perceptions. *Health Services Research* 1976;11:396–415.
- Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BS and Morris JR. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *International Journal of Health Services* 1976;6:393–415.
- Ware JE Jr and Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;30:473–83.
- Brooks RG. EuroQoL – the current state of play. *Health Policy* 1996;37:53–72.
- Horsman J, Furlong W, Feeny D and Torrance G. The Health Utility Index (HUI): concepts, measurement properties and applications. *Health and Quality of Life Outcomes* 2003;1:54.
- Acquadro C, Berzon R, Dubois D et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the patient-reported outcomes (PRO) harmonization group meeting at the Food and Drug Administration, February 16, 2001. *Value in Health* 2003;6:522–31.
- Stull DE, Leidy NK, Parasuraman P and Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Current Medical Research and Opinion* 2009;25:929–42.
- Rosser RM and Watts VC. The measurement of hospital output. *International Journal of Epidemiology* 1972;1:361–8.
- Rosser R and Benson T. New tools for evaluation: their application to computers. In: Anderson J (ed) *Medical Informatics Europe 78, First Congress of the European Federation for Medical Informatics: proceedings*. Cambridge, England, September 4–8, 1978 (lecture notes in medical informatics). Berlin: Springer Verlag, 1978, pp. 701–10.
- Rosser R, Allison R, Butler C, Cottee M, Rabin R and Selai C. The Index of Health-Related Quality of Life (IHQL): a new tool for audit and cost-per-QALY analysis. In: Walker SR and Rosser RM (eds) *Quality of Life Assessment: key issues in the 1990s*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1992, pp. 179–84 and 455–66.
- Bakken S, Cimino JJ, Haskell R et al. Evaluation of the Clinical LOINC (Logical Observation Identifiers, Names, and Codes) Semantic Structure as a Terminology Model for Standardized Assessment Measures. *Journal of the American Medical Informatics Association* 2000;6:529–38.
- Chisholm J. The Read Clinical Classification: the NHS has acquired a coding system designed for the computer age. *British Medical Journal* 1990;300:1092.
- NHS Information Authority. *Clinical Terms Version 3 (the Read Codes): introduction and overview*. 2000. www.connectingforhealth.nhs.uk/systemsandservices/data/readcodes/docs/chap1.pdf
- Benson T. *Principles of Health Interoperability HL7 and SNOMED*. London: Springer, 2010.
- Ware JE, Kosinski M and Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care* 1996;34:220–33.
- Messick S. Validity of psychological assessment. *American Psychologist* 1995;50:741–9.
- Clark LA and Watson D. Constructing validity: basic issues in objective scale development. *Psychological Assessment* 1995;7:309–19.
- Streiner D and Norman G. *Health Measurement Scales: a practical guide to their development and use (4e)*. Oxford: Oxford University Press, 2008.
- Netemeyer RG, Bearden WO and Sharma S. *Scaling Procedures: issues and applications*. Thousand Oaks, CA: Sage Publications, 2003.
- Fayers PM and Machin D. *Quality of Life: the assessment, analysis and interpretation of patient-reported outcomes*. Chichester, UK: Wiley, 2007.

- 31 Finster M and Wood M. The Apgar Score has survived the test of time. *Anaesthesiology* 2005;102:855–7.
- 32 Teasdale G and Jennett B. Assessment of coma and impaired consciousness: a practical scale. *The Lancet* 1974;304:81–4.
- 33 Murray DW, Fitzpatrick R, Rogers K *et al.* The use of the Oxford Hip and Knee Scores. *Journal of Bone and Joint Surgery – British Volume* 2007;89-B:1010–14.
- 34 Higginson IJ and Carr AJ. Using quality of life measures in the clinical setting. *British Medical Journal* 2001;322:1297–300.
- 35 Sederer LI, Dickey B and Eisen SV. Assessing outcomes in clinical practice. *Psychiatric Quarterly* 1997;68:311–25.
- 36 Donaldson MS. Taking PROs and patient-centered care seriously: incremental and disruptive ideas for incorporating PROs in oncology practice. *Quality of Life Research* 2008;17:1323–30.
- 37 Rose M and Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. *Quality of Life Research* 2009;18:125–36.
- 38 Hawthorne G. Assessing utility where short measures are required: development of the Short Assessment of Quality of Life-8 (AQoL-8) instrument. *Value in Health* 2009;12:6, 948–57.
- 39 Paz SH, Jiu H, Fongwa MN, Morales LS and Hays RD. Readability estimates for commonly used health-related quality of life surveys. *Quality of Life Research* 2009;18:889–900.
- 40 Calderon JL and Beltran R. Pitfalls in health communication: healthcare policy, institution, structure and process. *Medscape General Medicine* 2004;6:e7.
- 41 McDowell I. *Measuring Health: a guide to rating scales and questionnaires* (3e). Oxford: Oxford University Press, 2006.
- 42 Sturgis P, Thomas R, Purdon S, Bridgwood A and Dodd T. *Comparative Review and Assessment of Key Health State Measures of the General Population*. London: Department of Health, 2001.
- 43 Fitzpatrick R, Bowling A, Gibbons E *et al.* *A Structured Review of Patient-Reported Measures in Relation to Selected Chronic Conditions, Perceptions of Quality of Care and Carer Impact*. Patient-Reported Health Instruments Group. Oxford: Report for the Department of Health, 2006. phi.uhce.ox.ac.uk/pdf/ChronicConditions/dh_reviews_november_2007.pdf
- 44 Browne J *et al.* *Patient Reported Outcome Measures (PROMs) in Elective Surgery*. London: Report for the Department of Health, 2007. www.lshtm.ac.uk/hsru/research/PROMs-Report-12-Dec-07.pdf
- 45 Johnson JA and Pickard AS. Comparison of EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care* 2000;38:115–21.
- 46 Hawthorne G, Richardson J and Atherton Day N. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals of Medicine* 2001;33:358–70.
- 47 Larson CO, Hays RD and Nelson EC. Do the pictures influence the scores on the Dartmouth COOP Charts. *Quality of Life Research* 1992;1:247–9.

FUNDING

No external funding received.

CONFLICTS OF INTEREST

TB is the owner of Abies Ltd, which has the IPR on *howRu*.

ADDRESS FOR CORRESPONDENCE

Tim Benson
Abies Ltd
14 Pinewood Crescent
Hermitage
Thatcham RG18 9WL
UK
Tel: +44 (0)1635 203162
Email: tim.benson@abies.co.uk

Accepted May 2010

If you wish to use *howRu* please contact Tim Benson at Abies Ltd.
Those wishing to use the SF-12 should contact the QualityMetric Inc. (www.qualitymetric.com).

