



Editor's choice
Scan to access more
free content

A systematic review of measurement tools of health and well-being for evaluating community-based interventions

Mithilesh Dronavalli, Sandra C Thompson

Western Australian Centre of Rural Health, University of Western Australia, Geraldton, Western Australia, Australia

Correspondence to

Dr Mithilesh Dronavalli,
Western Australian Centre of Rural Health, University of Western Australia, 167 Fitzgerald St, Geraldton, WA 6530, Australia;
dr.mit@me.com

Received 13 January 2015

Revised 7 March 2015

Accepted 10 May 2015

Published Online First

3 June 2015

ABSTRACT

Background Those interested in evaluating the effectiveness of community interventions on health and well-being need information about what tools are available and best suited to measure improvements that could be attributed to the intervention.

This study evaluated published measurement tools of health and well-being that have the potential to be used before and after an intervention.

Methods A literature search of health and sociological databases was undertaken for articles that utilised measurement tools in community settings to measure overall health, well-being or quality of life. Articles were considered potentially relevant because they included use of measurement tools related to general health or well-being. These tools were evaluated by further searching of the literature to assess each tool's properties including: reliability; validity; responsiveness; length; use in cross-cultural settings; global health or well-being assessment; use of subjective measures; clarity and cost. A composite score was made based on the average rating of all fields.

Results Of 958 abstracts that were screened, 123 articles were extracted for review. From those articles, 27 measurement tools were selected and assessed. Based on the composite score assessing across all domains, five tools were rated as excellent.

Conclusions While tools may need to be selected for particular aims and interventions, a range of potential well-described tools already exist and should be considered for use in preference to ad hoc or bespoke tools. Any of the five tools rated as excellent are recommended to assess the impact of a community intervention.

AIM

To systematically review and evaluate the characteristics of measurement tools that measure community health and well-being.

BACKGROUND

Often in public health, interventions are proposed or implemented with a community or group of individuals to improve their health and well-being. Those with an interest in evaluation want information about what tools are available and those that are best suited to measure improvements that could be attributed to the intervention. There are obvious benefits from using efficient and standardised tools of measurement and for which population norms are available.

This study evaluates standardised instruments (tools) that measure community health and well-being, and could potentially be used before and

after an intervention. Characteristics of the tools such as reliability, validity, responsiveness and other key features were assessed.

METHODS

The methods for this article involved a two-stage process.

The first stage was identifying articles that reported on health and/or well-being in the general population and which used or reported on a measurement tool as a part of assessing health and well-being. In this study, the term general population refers to adults over the age of 18. This definition also excludes studies that focused only on the elderly (eg, only adults >70 years of age).

A literature search was undertaken on or before 9 December 2014 on the following health and sociological databases—ERIC, JSTOR, Proquest, Soc Index, Web of Science, Psych Info, EMBASE, Wiley Online Library, Medline, Cochrane, Informit, Cinahl Plus and Project Muse. While search terms were slightly modified for each database, the generic search terminology was based on the following terms: 1. Community AND 2. 'Overall Health' AND 3. (Wellbeing or 'Quality of Life') AND 4. (assessment* or questionnaire* or interview* or rating* or scale* or measure* or test* or survey* or instrument*).

The search identified 958 abstracts, which were initially screened by review of the title and abstract. Articles were considered potentially relevant if published in or after 1990 and if they included the use of measurement tools related to physical health or well-being (emotional, mental, social, etc) in the general population. Tools published before 1990 and not subsequently reported on were not included.

From the 123 articles, 81 potential tools used to assess health and/or well-being were identified. An additional 15 tools were located through a Google search. Of the total 96 potential tools, the following criteria were applied. All tools had to:

- ▶ Be named and used in multiple studies (23 tools excluded);
- ▶ Be focused on health or well-being (17 tools excluded; eg, not focused on crime, poverty, environment, etc);
- ▶ Have a subjective component (9 tools excluded);
- ▶ Have psychometric data (7 tools excluded);
- ▶ Be globally relevant, that is, not just about one country, culture or locality (6 tools excluded);
- ▶ Have data on well-being (4 tools excluded);
- ▶ Have more than one question (3 tools excluded).



CrossMark

To cite: Dronavalli M, Thompson SC. *J Epidemiol Community Health* 2015;69:805–815.

The second stage was an in-depth evaluation of the 27 tools. The name of each tool was pasted into Google and into One Search, the single point search tool by the library of the University of Western Australia. Tools were evaluated based on a hierarchy of evidence. This was undertaken sequentially, starting with systematic reviews of tools, then with information from the tool's own website, followed by psychometric articles of the tools reported in journal articles, summary sites of tools (eg Rehab Measures and Par-qol) and finally with original articles using the tool. Additional searching was undertaken to collect information on costs of using each tool. Most of the information extracted for each tool came from systematic reviews or the tool's own website or main psychometric articles making the data accurate. Manual searching was not used to obtain further references due to the large number of articles located.

For each tool, the reliability, validity, responsiveness, the availability of population norms, length, clarity of questions, cross-cultural use, cost and domains measured were assessed. Also assessed was whether the tool measured health and well-being, used subjective measures or whether the tool included a global assessment of either health or well-being. All of these tools were suitable for the general population and therefore relevant to community interventions.

Reliability of each identified instrument was assessed through test-retest measures, and a measure of internal consistency (Cronbach's α). Validity was assessed based on standard definitions for various types of validity (see [table 1](#)). Responsiveness was assessed based on the timeframe of inquiry listed in a tool, that is, whether it referred to the present or near past, so that re-testing allowed the potential for change in health and well-being to be assessed.

Since it is desirable for a tool to be used in a population that is often culturally heterogeneous, information on the tool in terms of its assessment and use in cross-cultural settings was also assessed.

The number of items within the tool and in some cases the time to assess participants was documented, as well as the domains it covered.

The tool was assessed as to whether it included a global measure of health or well-being (eg, "How do you feel about your current general health?" and "Do you feel happy?").

Table 1 Types of validity and their definition

Type of validity	Definition
Face validity	Tool measures content according to the lay-person
Content validity	Tool measures content according to an expert based on theory
Criterion validity	The criteria of the tool match other tools
Concurrent or Convergent validity	Tool gives similar scores as other tools on the same subjects
Discriminant validity	Tool gives high scores for well people and low scores for sick people
Predictive validity	Tool predicts significant differences among different groups
Construct validity	Tool measures what it is supposed to
Internal validity	Measurement results warrant a causal conclusion, that is, observed changes can be attributed to the programme or intervention
External validity	The measurement results of a tool can be generalisable to the whole population
Multidimensionality	Subfactors of the tool have low intercorrelations
Divergent validity	The tool asks questions only related to the construct

Global questions of health and well-being are required as they are a summary measure of the person's state of well-being and enable the tool's overall score to be compared with the response to the global question. This comparison can be used to gauge construct and divergent validity. Discriminant validity can also be tested using this mechanism, as a high global question score should reflect a high overall score and vice-versa.

The clarity of the questions was also assessed. All the questionnaires were read by the first author and each tool's questionnaires were classified as being either 'easy', 'moderate' or 'complex' to understand. The clarity of the top five tools was then re-evaluated by the second author. For a tool to have better clarity, a lay person should be able to read, understand and respond to the questions with ease. There should be minimal ambiguity or use of idioms or phrases not understood by the general population (especially people from linguistically diverse groups). There should not be too many conditions in the statement as this would require complicated thought processes to answer accurately. Also, tools were classified as 'complex' if the tool documentation or reviews of the tool stated that the tool required extensive interviewer training.

An example of a question from an easy tool is "How satisfied are you with your standard of living?" (from the Personal Wellbeing Index). An example of a question from a tool rated as of medium clarity is from the LSIA when testing resolution and fortitude. The description of the question is: "The extent to which R accepts personal responsibility for his life; the opposite of feeling resigned, or of merely condoning or passively accepting that which life has brought him". The description of an answer with a 5 rating (the highest score) is: "Try and try again attitude. Bloody but unbowed. Fights back; withstanding, not giving up. Active personal responsibility—take the bad and the good and make the most of it. Wouldn't change the past".

An example of a question from a tool rated as complex is from Health Utilities Index-3. The question related to hearing. Here is the answer for 2 out of 6 points on the Likert scale: "Able to hear what is said in a conversation with one other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least three other people."

Documented data about each tool was referenced.

Finally, the standardised tools were evaluated for likely costs associated with their use. This is an important matter for any community study where costs associated with purchase of standardised tools may make of an otherwise excellent tool unaffordable.

Of the 27 measurement tools identified, a search of the relevant literature was conducted for each tool, separately yielding 51 articles or entries on the web.

- I. Four systematic review articles of tools.
- II. Three tools had information from the tool's own website.
- III. Twenty-four articles assessing the psychometric properties of the tool(s).
- IV. Three tools had entries on summary references of tools (eg Rehab Measures, Par-qol, Corsini Encyclopaedia of Psychology).
- V. Two original articles using the tool.
- VI. Fourteen entries were related to cost.

Scoring of each criterion was used in conjunction with colour-coding (green for 'high quality', yellow for 'average quality' or red for 'low quality'), to assist with an overall assessment of each tool, and ready identification of any weaknesses and strengths on the selected tool properties. [Table 2](#) contains the key to the colour classifications for the relevant variables.

Table 2 Key to colour coding/scoring of tool properties

Variable	Green (+1 point)	Yellow (+0.5 points)	Red (−1 point)
Test–retest (use middle of interval)	ICC or correlation >0.8. κ >0.7	ICC or correlation >0.6 or κ >0.5	ICC or correlation <0.6 or κ <0.5
Validity	Good validity on a number of factors	–	Ceiling or floor effects with poor discriminant validity
Responsiveness	Set in the present or past few weeks, known to be responsive	–	Known to be not responsive, set in the past or on the life continuum, floor or ceiling effects
Measure of health	Yes	–	No
Measure of wellbeing or quality of life	Yes	–	No
Clarity	Easy	Moderate	Complex
Use of subjective measures	Yes	–	No
Global assessment of health or wellbeing made	Yes	–	No
Used cross-culturally	Many languages	Other than English	Only English, not documented
Length	≤10 min or ≤15 items (either)	<30 mins or 50 items (either)	≥50 items ≥30 min
Cronbach's α (use middle of interval)	≥0.8	≥0.6 to <0.8	<0.6
Cost	Free to use for research purposes	Reasonable one-off payment	Expensive

ICC, intraclass correlation coefficient.

From scoring of individual properties, tools were colour coded and a composite score was determined. A green entry meant the criteria for each domain scored 1 point, a yellow entry scored 0.5 points and a red entry was penalised 1 point. Missing entries were not counted and also not included in the denominator.

The composite scores were displayed graphically and summarised into four categories based on the respective cut-offs of the composite score (Poor <0.5; Mediocre 0.5–0.75; Good >0.75–0.85; Excellent >0.85). By definition, the score could have a maximum of 1 and minimum of −1.

RESULTS

From reviewing 123 articles that described measurement tools of health and well-being, and searching Google, we identified 96 tools. After applying our selection criteria, 27 instruments measuring health and/or well-being were identified for closer analysis of their psychometric and other properties. All 27 tools identified by the search criteria (figure 1) were scored and all were suitable for the general population, although three tools were primarily aimed at assessing patients with a particular disease state and two tools had been used for studies in the elderly. Six of the tools were assessed as not measuring health but, rather, measuring well-being. Although the mHAQ does not measure well-being, this tool contributed information beyond tools that only measured physical health, as the mHAQ measures ability to undertake activities of daily living, which heavily influence well-being.

Table 3 lists all the tools in rank order based on their composite score with respective colour coded data. For most of the instruments assessed, there were data on most of the properties. However, for seven tools there was no assessment of test–retest reliability. For one tool, there was no English version of the instrument, precluding assessment of the clarity of the questions.

Four tools did not have a reported Cronbach α to assess internal consistency; for a further three tools the Cronbach α was not mathematically relevant due to the way these tools are constructed.

The Cronbach α is designed in such a way that it assumes items in a measurement tool have equal SDs and are equally

correlated. While this is not a requirement for measuring reliability, it is a requirement for the Cronbach α . Alternative measures such as the Tarkkonen ρ have been presented but are not widely used.

There was no information available on the cost of use for one tool. Overall, complete scoring on the predetermined criteria was possible for most instruments, so table 3 is substantially populated and largely complete.

Of the 27 tools, 25 can be self-administered. The Behavioural Risk Factor Surveillance System (BRFSS) is a telephone survey administered nationally in the USA and is not self-administered. The Quality of Wellbeing (QWB) scale currently requires extensive training of the interviewer. Although there is a simpler self-administered version, it has not yet been fully reviewed.

The composite score was classified into four classes. Five tools with a composite score above 0.85 were classified as excellent. Nine tools were classified as good based on a composite score between >0.75 and 0.85 inclusive. Seven tools were classified as mediocre with a composite score between 0.5 and 0.75. Six tools were classified as poor and they had a composite score below 0.5. The median composite score was 0.77.

The five tools that were rated as excellent are the Quality of Life Scale (QOLS), Personal Wellbeing Index (PWI), Community Wellbeing Index (CWI), the WHO Quality of Life—Brief (WHOQOL-BREF) and the Health Related Quality of Life from the Dartmouth Co-operative Information Project/World Organisation of National Colleges, Academies and Academic Associations of General Practices/Family Physicians Charts (HRQOL from COOP/WONCA Charts).

Figure 2 represents the composite score of each tool graphically in a bar graph.

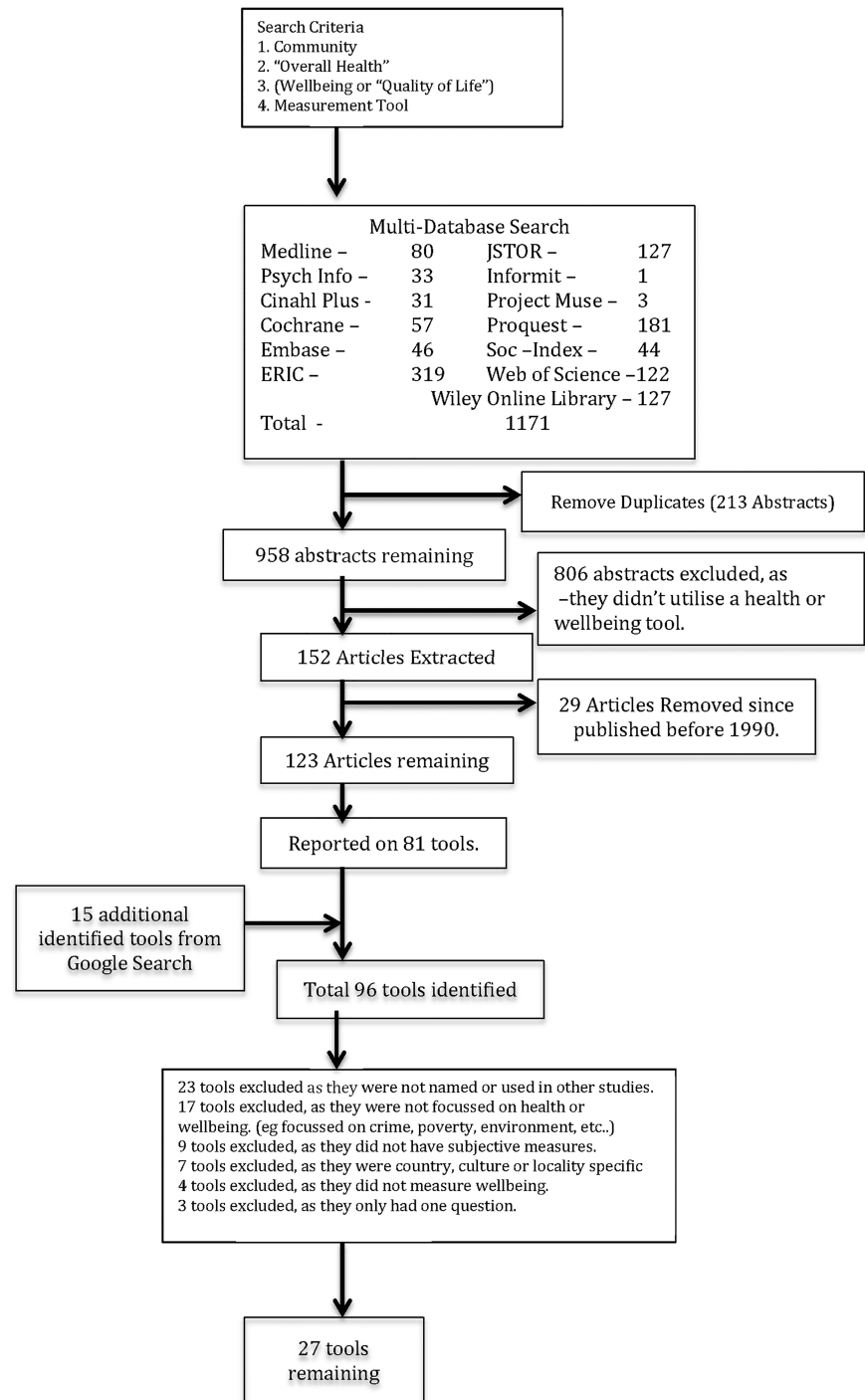
Table 4 is a summary table that reports the number of tools in each class with the composite score cut-offs discussed.

DISCUSSION

Those with a desire to evaluate their community interventions are presented with a selection of measurement tools of health and well-being. Some of these tools are excellent for this purpose.

Of the top five tools classified as excellent, four are known to be free; the HRQOL from the COOP/WONCA chart requires a

Figure 1 Flow chart of literature search for measurement tools and their evaluation.



one-time payment of \$15. This means that all five of these tools are affordable to use in community studies. Note that the CWI only had a Spanish version. Most of the top tools are quite short and are easy to administer with good clarity. By definition, the top tools have been used in cross-cultural settings, have good reliability and validity and are based in the present for good responsiveness. Furthermore, all tools in [table 3](#) have a list of domains they cover and this will be useful for investigators who have an interest in certain domains when planning studies.

There are many benefit of using standardised tools to measure the effectiveness of community interventions. There seems little justification for developing new (unvalidated) tools when assessing an intervention, as there are excellent standardised tools that are either free or low cost to use.

All of the top five tools rated as excellent come from internationally recognised sources. The PWI and CWI groups are affiliated with The International Wellbeing Group and present a large body of literature regarding the Quality of Life and Wellbeing subject. WHOQOL-BREF is the brief version of the WHOQOL, which has 100 items. The WHOQOL-BREF has 26 items and is psychometrically representative of the larger WHOQOL instrument, and although there is some loss in internal consistency it ranks higher because of its reduced administrative burden. The HRQOL is derived from assessing the health and well-being sections of the COOP/WONCA charts. HRQOL is exciting as it uses pictures to assess health and well-being, which makes it appropriate for use with participants with low literacy levels.

Table 3 Measurement tools and their properties

Measurement tool	Cronbach's α	Test-retest	Validity	Population norms	Responsiveness	Measure of health	Measure of well-being	Subjective measure used	Global assessment made	Length	Clarity of questions	Cross-cultural use	Cost	Domains
QOLS-Quality of Life Scale ¹	0.82–0.92	0.78–0.84	Good convergent and discriminant validity	Yes	Responsive to intervention, for example, treatment in chronic disease	1	1	1	1	16 items, 5 min	Easy,	Many languages	Free for research and clinical use ¹	Material and physical well-being, relationships with other people, social, community and civic activities, personal development and fulfilment and recreation
PWI-Personal Wellbeing Index ²	0.7–0.85	ICC=0.84	Good construct and convergent validity	Yes	Set in present	1	1	1	1	7 items	Easy	Many languages	Free for research use. (From email to author)	Living standards, Health, Achieving in life, Relationships, Safety, Community connection and Future security
CWI-Community Wellbeing Index ³	PSI=0.82–0.85	Unknown	Good internal, external and discriminant validity	Yes	Set in present	1	1	1	1	10 items	Questionnaire in Spanish	Used in Spain	Free (From email to author)	Community services, Community attachment, physical and social environment
WHOQOL-BREF ^{4 5}	0.68–0.82 all domains	ICC=0.85–0.95	Good criterion validity with WHOQOL, convergent validity, content, construct and divergent validity. Good predictive validity for depression	Yes	Set in present.	1	1	1	1	26 items	Easy	Many languages	Free for research use ⁶	Physical, psychological, Social and environment
HRQOL from COOP/WONCA charts ⁷	0.766	>0.75	Good predictive validity with age and physical fitness. Good convergent validity with SF-36	Yes	Past 2 weeks	1	1	1	1	<5 min	Easy, uses pictures	Many languages	\$15 One off payment ⁸	Physical fitness, Feelings, Social activities, Overall health and pain
WHOQOL-100 ^{9 10}	0.967	Highly correlated over 2–8 weeks	Good discriminant validity between sick and healthy. Good content, construct, divergent and convergent validity	Yes	Past 2 weeks	1	1	1	1	100 items	Easy	Many languages	Free unless pharma or commercially funded, then it is 350 pounds ¹¹	Overall health and quality of life, Pain and discomfort, Energy and fatigue, Sleep and rest, Positive feelings, thinking, self-esteem, appearance, negative feelings, mobility, ADL, dependence on medicines, working capacity, social support, sexual activity, physical safety and security, home environment, financial resources, health and social care, opportunities, participation, environment, transport and Beliefs
SIP-Sickness Impact Profile ¹²	0.9–0.92	0.88	SIP68 Good criterion validity with SIP. Good content validity	Yes	Can detect change, Responsiveness index: SIP68=0.62 SIP136=0.64	1	1	1	1	SIP: 136 items, 20–30 min SIP68: 68 items 15–20 min	Easy	Many languages	Free to use ¹³	Physical, Mental and Social
SF12-Short Form 12 ¹³	Not listed	0.89–0.76	Good criterion validity with SF36, discriminant validity (slightly lower than SF36)	Yes	Past 4 weeks	1	1	1	1	2 min or 12 questions	Easy	Many languages	Cost proportional to size of study ¹⁴	Physical and mental health

Continued

Table 3 Continued

Measurement tool	Cronbach's α	Test-retest	Validity	Population norms	Responsiveness	Measure of health	Measure of well-being	Subjective measure used	Global assessment made	Length	Clarity of questions	Cross-cultural use	Cost	Domains
BRFSS-Behavioural Risk Factor Surveillance System ¹⁵	Not listed	High for Health and Wellbeing domains	Good convergent validity	Yes	Used as a longitudinal measure	1	1	1	1	Not all domains necessary	Easy. Telephone Survey	Not much data	Free to use. (From email to author)	1. Access to healthcare/general health 2. Immunisation, preventive screening, and testing 3. Physical activity measures 4. Chronic disease 5. Mental health measures 6. Overweight and obesity measures 7. Tobacco and alcohol use measures 8. Responsible sexual behaviour measures 9. Injury risk and violence
SOC13-Sense of Coherence ¹⁶	0.91 (8 published studies)	0.91 over 2 weeks	Good construct, consensual and discriminant validity. Good criterion validity to a number of studies	Yes	Set in present	1	1	1	0	10–15 min 13 items	Easy	Many languages	Free for academic and non-commercial purposes ¹⁷	Global orientation to oneself and one's environment (19 r's); stressors (11 r's); health, illness and well-being (32 r's); attitudes and behaviour (5 r's)
SOC29-Sense of Coherence ¹⁶	0.82 (5 published studies)	0.91 over 2 weeks	Good construct, consensual and discriminant validity. Good criterion validity to a number of studies	Yes	Set in present	1	1	1	0	15–20 min 29 items	Easy	Many languages	Free for academic and non-commercial purposes ¹⁷	Global orientation to one and one self's environment, stressors, health, illness, well-being, attitudes and behaviours
WBS-Wellness Belief Scale ¹⁸	0.88–0.94	Unknown	Good discriminant and predictive validity	Healthy women <25	Set in present	1	1	1	1	20 items or equivalent 9 items	Easy	Not documented	Requested	Health behaviours, Health beliefs and Response to illness
RAND12 ^{19–21}	Not measured	Unknown	Good construct validity. Better discriminant validity than SF12	Yes	Not as responsive as disease specific scales for heart failure	1	1	1	1	2 min or 12 questions	Easy	Many languages	Cost proportional to size of study. Uses the SF-12 Questionnaire ²²	Physical and Mental dimensions
SF36-Short Form 36 ²²	0.9	ICC=0.81 General Health ICC=0.52 Mental Health	Good content, concurrent, criterion, construct and predictive validity	Yes	Based on the past 1-week or 4 weeks	1	1	1	1	36 items	Easy	Many languages	Cost of use proportional to size of study ¹³	Physical, social, role physical, role emotional, mental health, energy and pain
SWLS-Satisfaction With Life Survey ²⁴	0.87	0.82	Good discriminant validity and construct validity	Yes	Speaks of life in general, does not refer to immediate past. Poor responsiveness	0	1	1	1	5 items	Easy	Many languages	Free to use ²⁵	Life in general
HPLP-II-Health-Promoting Lifestyle Profile ²⁶	0.94	Unknown	Good construct validity	Yes	Set in Present	1	1	1	0	52 items	Easy	Many languages	Free for non-profit. Needs permission to use in a study ²⁷	Health responsibility, Physical Activity, Nutrition, Spiritual growth, interpersonal relationships and Stress managements
EQ5D-EuroQol ²⁸	Not mentioned	ICC 0.85 3 week retest	Moderate convergent validity with SF 12. Good predictive validity for recent illness	Yes	Set in the present. Poor sensitivity due to ceiling effect	1	1	1	1	1 min	Easy	Many languages	Cost is proportional to size of study ¹⁹	Mobility, Self-care, Usual activity pain/discomfort and anxiety/depression

Continued

Table 3 Continued

Measurement tool	Cronbach's α	Test-retest	Validity	Population norms	Responsiveness	Measure of health	Measure of well-being	Subjective measure used	Global assessment made	Length	Clarity of questions	Cross-cultural use	Cost	Domains
QWB-Quality of Wellbeing Scale ²⁸	Not appropriate	0.9 over a few days	Good convergent validity with SF 20 and COOP. Good construct and predictive validity	Yes	Uses the past 3 days. Responsive to treatment interventions	1	1	1	0	10–30 min	Complex	10 different languages	Free for research and not for profits ⁴⁰	Mobility, Physical activity, Social activity and symptoms/problems
WEL-Wellness Evaluation of Lifestyle ^{31 32}	0.61–0.89	0.88	Good concurrent validity with Testwell and other scales. Good face validity	Yes	Responsive to intervention in a 4 week RCT	1	1	1	1	134 items, 12 subscales	Moderate	Culturally adapted to Korean, Hebrew and Turkish and different age groups	\$360 for 500 patients ⁴¹	Spirituality, Self-regulation, Work and leisure, Friendship, Love, Total self-regulation, Perceived wellness and Total wellness
SEQOL-Self-evaluation of Quality of Life ³⁴	0.75	>0.8	Good concurrent validity with NHP and SCREENQOL. Acceptable construct validity, external reliability	Yes	111 respondents needed to detect a 3% difference	0	1	1	1	317 items	Easy	Not Determined	Free for research purposes. (From email to author)	Well-being, satisfaction, Happiness, Family work and Leisure, Fulfilment of needs, Satisfaction with relationships, Realising life potential, Objective factors and Overall QOL
SCL-90-R-Symptom Checklist 90 R ^{35–37}	0.97	Unknown	Good concurrent, construct validity and convergence validity	Yes	Set in present	0	1	1	1	90 items, 12–15 items	6th Grade Level	Many languages	50 patients \$121 ³⁸	Somatisation (SOM), Obsessive-Compulsive (OBS), Interpersonal Sensitivity (INT), Depression (DEP), Anxiety (ANX), Hostility (HOS), Phobic Anxiety (PHOB), Paranoid Ideation (PAR), and Psychoticism (PSY). The global measures are referred to as the Global Severity Index (GSI), the Positive Symptom Distress Index (PSDI), and the Positive Symptom Total (PST)
FS-Friendship Scale ³⁹	0.81	Unknown	Good concurrent validity with WHOQOL-Bref, AQoL, SF-12MCS and PCS. Good discriminant validity	Older Adults	Set in the present	0	1	1	0	6 items	Easy	Not documented	Author passed away, likely free	Friendship
HUI-3-Health utilities index ³⁸	Not appropriate	Moderate or better κ for 8/10 questions	Good face and predictive validity	Yes	Set in the present	1	1	1	0	5–10 min	Complex	Many languages	\$5000 for HUI3 ³⁹	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition and Pain
NHP-Nottingham Health Profile ^{28 41 42}	0.72	0.93	Poor discriminant validity, 50% answered no to all 38 questions. Poor multidimensionality (high-factor intercorrelations)	Yes	Set in the present, but mode is zero so likely has poor responsiveness	1	1	1	0	38 items, 5–10 min	Easy	Many languages	Not Free ⁴⁰	Emotional, Social and Physical

Continued

Table 3 Continued

Measurement tool	Cronbach's α	Test-retest	Validity	Population norms	Responsiveness	Measure of health	Measure of well-being	Subjective measure used	Global assessment made	Length	Clarity of questions	Cross-cultural use	Cost	Domains
mHAQ-Modified Health Assessment Questionnaire ⁴⁴	0.87	Unknown	Many patients scored 0. Poor discriminant validity. Good construct validity. Good convergent validity with WHODAS	No. Mainly used in arthritic patients	Set in present	1	0	1	0	8 items	Easy	Many languages	Free to use ⁴⁵	Activities of Daily Living
LSIA-Life Satisfaction Inventory A ^{46 47}	0.76	ICC=0.87	Good concurrent validity. Poor multidimensionality (high-factor intercorrelations)	Yes	Not responsive, looks back onto life	0	1	1	0	20 items —10 min	Moderate	Only English	Free to use	Zest, Resolution, Congruence and Self-concept
SEIQOL-DW-Schedule for the evaluation of individual quality of life —Direct Weighting ^{48–51}	Not appropriate	$\kappa=0.51$ after 1 week retest correlations in stable patients (Moderate)	Good construct, convergent, discriminant validity for well-being, moderate to poor discriminant validity for health	Not Suitable	Poor responsiveness to interventions. Further research needed	0	1	1	0	30 min	Moderate	Many languages	Free for research purposes ⁵²	Physical, Emotional and Family

Each property for every tool was assessed for quality using the criteria in Table 2. If a property for a tool was 'high quality', it was scored +1 and the cell for that property was coloured green. For properties of tools that were of 'average quality' the cell was scored +0.5 and coloured yellow. For properties of tools that were 'low quality', cells were scored -1 and coloured red. The composite score was the summation of scores for all non-missing properties divided by the number of non-missing properties available for that tool. Tools are ranked from highest to lowest in this table based on their composite scores.

ICC, intraclass correlation coefficient.

The QOLS scale was constructed by John Flanagan in the 1970s and has been cross-culturally adapted with relevant psychometric assessments. The PWI tool was constructed by an international collaboration headed by Professor Cummins of Deakin University in Australia. The CWI tool arose from a collaboration organised by Professor Forjaz of the National School of Public Health in Madrid, Spain. The UK version of the WHOQOL and the WHOQOL-BREF were developed by a team commissioned by WHO headed by Professor Skevington of the University of Bath, UK. The HRQOL and COOP/WONCA charts were constructed as a part of international collaboration based in Dartmouth Medical School in USA.

The 16 item QOLS tool by Burckhardt and Flanagan essentially describes the ideal enriched life according to a certain ideal standard. That is, being happily married with children, in a fulfilling job, engaging with the community, having good opportunities for recreation, with material comforts and having good friends. The QOLS tool scores people by how closely they fit this ideal life. While many may agree that this is the ideal life-style, not everyone follows this ideal. Some people are single without kids, some value career above community engagement, relationships and recreation. Therefore, this tool likely reflects the aspirations of the majority and incorporates the dimensions that conventional wisdom has shown are important determinants of health and well-being. Also the QOLS tool has been widely used in people with a wide variety of chronic diseases including diabetes, osteoarthritis, gastrointestinal disorders, rheumatoid arthritis, chronic obstructive pulmonary disease, heart disease, lower back pain, post-traumatic stress disorder and other chronic diseases.¹

The PWI by Cummins *et al* is a short tool, with only seven items. While being brief, it adequately balances health, well-being, relationships and community connectedness. Since the PWI is so brief, the discriminatory nature of the tool is limited. Advantages are that it measures future security and, also, it is very easy to administer. The PWI has been used in many studies that require assessment of general well-being. It is not generally used for specific disease states, but more often is used to compare healthy subgroups with regard to their well-being, for example, adolescents, or certain communities or countries (Australia, China, Macau). The PWI has also been used in assessment of well-being of various psychological states such as depression.²

The CWI by Forjaz is purely a community connectedness tool. It focuses on the fit of the individual with the surrounding community. It purposely does not measure individual characteristics of health and well-being but focuses on health and well-being from a community perspective. For example, "are you satisfied with the health services of your town or city", rather than a person's overall health. In every question of the 10 items, the CWI relates to the town or city of the resident. This tool is therefore particularly useful to assess a community as a whole rather than a collection of individuals, and is an important tool to use when evaluating community interventions. The CWI tool was developed only very recently, and is still being translated into and tested in English. It has only been used in a few studies in Spain, with one such study being among the elderly.

The HRQOL tool by Nelson *et al* is an ingenious tool that uses meaningful pictures attached to a normal Likert scale of answers for each question, making it suitable for low literacy respondents. Each question has five responses with ordered pictures for each severity. An interesting research question is whether responses are more standardised when pictures are attached to each Likert scale. The HRQOL is quite brief and

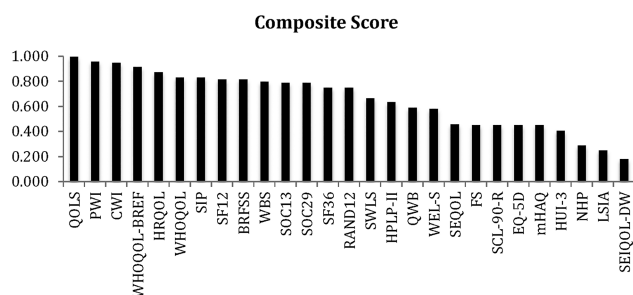


Figure 2 All measurement tools with their respective composite scores.

very easy to administer because of the pictures. It is very much focused on the individual rather than the community, with only one in six questions related to community connectedness. The HRQOL is also a very general tool, similar to the PWI, and so the discriminatory nature of the tool is compromised. The HRQOL has mainly been used to assess the general health and well-being of patients in chronic disease states including diabetes, chronic kidney disease, stroke and multiple sclerosis. Interestingly, the HRQOL has also been used in patients from China in the primary care setting where the pictures in the HRQOL may have been useful.

The WHOQOL-Bref has 26 items and, despite being termed brief, it is the longest and the most widely used of the top five tools. The WHOQOL-Bref has exceptional discriminating qualities as it is quite detailed. It uniquely measures the positive and negative attributes of physical and psychological health. The WHOQOL-Bref makes a detailed assessment of the individual and their role in the community. Since this tool is more detailed it may take longer to administer. The length of the tool may also affect responder comprehension. The WHOQOL-Bref has been employed internationally and is used for making comparisons between populations. There are over a 1000 studies using the WHOQOL-Bref, with most of these studies using it to measure health and well-being in populations. Healthy groups such as medical students and youth have also been studied using the WHOQOL-Bref. It has been used infrequently for disease states such as opioid addiction and HIV.

In conclusion, for a detailed assessment, the extensively studied WHOQOL-Bref is ideal and the QOLS is also suitable, but less generalisable. The PWI and HRQOL are easy to administer and brief, but may not be as discriminating as the other tools. Also, there are many advantages with using the HRQOL because it is picture based. Pictures overcome language barriers, may make assessment more standardised and facilitate easy administration of the tool. The CWI is purely a measure of the local community and the individual's view of his/her surrounding community. The CWI may be useful for assessments of a community or interventions that have an effect at the community level.

Table 4 Classification of measurement tools by composite score

Class of tools	Composite score	Number of tools
Excellent	>0.85	5
Good	>0.75–0.85	9
Mediocre	0.5–0.75	7
Poor	<0.5	6

Furthermore, there are many types of interventions that would be suitable for measuring change in overall health and well-being from the recommended top five tools.

- ▶ A new treatment for chronic disease, for example, a drug, operation or allied health intervention for a physical or psychological comorbidity.
- ▶ Programmes promoting primary prevention through targeting risk factors such as weight control, smoking cessation, increasing physical activity, harm reduction from alcohol and drugs.
- ▶ Alterations to the community at a community level, for example, intervention by a local council, increasing jobs, recreation and sport avenues, improving housing or health services.
- ▶ Targeted interventions at otherwise healthy subgroups, for example, for youth: a new gym, new indoor recreation centre or sports programme.

Note that different tools would be more suitable for different interventions. The CWI would be useful for interventions at the community level. The HRQOL or QOLS would be more suitable for treatments of chronic disease. The WHOQOL-Bref may be more suitable for primary prevention of risk factors and the PWI may be more suitable for targeted interventions in subgroups. A limitation of this study arises from the cut-offs and scores applied for the properties of each tool. While other cut-offs could have been chosen, it is unlikely that the ranking would have changed much. Some variables by their very nature are subjective, such as clarity and responsiveness. Clarity has been discussed extensively in the Methods section.

Responsiveness was determined based on the authors' assessment of the potential for the assessment to reflect change over time. Some tools showed change after an intervention and this was noted. Some tools were not responsive to major interventions and this was also noted. For the remaining tools, they were marked favourably if they were set in the present or the last few weeks. Tools were penalised in assessment if they referred to the whole life-course rather than to recent previous events. A life-course strategy would be expected to be associated with tools that are less responsive to interventions.

Articles that assessed or reported the validity of the measurement tools were used as the basis for our summary to comment on the validity of individual health and well-being tools. The findings of these articles were taken at face value and entered into table 3. However, many of the articles did not use common terminology for validity and did not assess validity in the same way. Some tools had significant floor or ceiling effects leading to poor discriminant validity. For example, the Nottingham Health Profile (NHP) scored 0 for over 50% of people in a study of the general population, exemplifying a major issue for using a tool to evaluate an intervention in the general population.

Despite limitations identified in some tools, the complete table is provided for information. We merely note that scoring has some elements of subjectivity, and potential for some different scoring criteria could be used to give different total scores and rankings. However, we have used a logical framework to distinguish between tools that could be used to assess health and well-being in community interventions. The information provided in table 3 enables others to identify instruments having properties and measure constructs that are more aligned with their purpose.

CONCLUSION

Our analysis identified the relevant literature and assessed the properties across various domains relevant to health and well-

being. Many of the tools are well constructed psychometrically, and some are freely available, while others require payment. Five measurement tools were rated as excellent using the scoring methods that we adopted. Our tabulation of the different properties across 27 instruments makes it easier to select an appropriate tool for evaluating the effectiveness of a community intervention to improve health and well-being.

There is an advantage in using these existing and well-characterised tools rather than constructing original tools, given that the existing choice includes free tools with sound psychometric properties, established reliability and validity, ease of use and, often, established population norms.

What is already known on this subject?

Many standardised measurement tools are available that measure health and well-being to evaluate community interventions, yet some investigators continue to use ad hoc tools. Some small reviews of a few tools exist.

What does this study add?

This study is a systematic comparison of all the relevant measurement tools of health and well-being found in the literature. Twenty-seven tools were found and further investigated for various properties and an overall comparison was made in a standard manner. This study allows investigators to pick an effective and appropriate measurement tool to evaluate their next community intervention.

Contributors ST developed the idea, carried out extensive editing and gave general guidance. MD wrote the review, search terms, carried out the search, populated the figures and tables, reviewed each selected tool individually and carried out the reference work.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The data for this article is basically Table 3 found in the article.

REFERENCES

- Burckhardt C, Anderson K. The Quality of Life Scale (QOLS): reliability, validity, and utilization. *Health Qual Life Outcomes* 2003;1:60.
- Cummins R. Personal Wellbeing Index. 2013. [cited 2014 Dec 9]. <http://www.deakin.edu.au/research/acqol/iwbg/wellbeing-index/pwi-a-english.pdf>
- Forjaz MJ, Prieto-Flores ME, Ayala A, et al. Measurement properties of the Community Wellbeing Index in older adults. *Qual Life Res* 2011;20:733–43.
- Amir M, Fleck M, Herrman H, et al. Reliability, validity and reproducibility of the WHOQOL-BREF in six countries. *Qual Life Res* 2000;9:320.
- Skevington SM, Lotfy M, O'Connell KA, et al. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res* 2004;13:299–310.
- Cost of WHOQOL-Bref. [cited 2014 Dec 9]. <http://depts.washington.edu/seaqol/WHOQOL-BREF>
- Martín-Díaz F, Reig-Ferrer A, Ferrer-Cascales R. Assessment of health-related quality of life in chronic dialysis patients with the COOP/WONCA Charts. *Nephron Clin Pract* 2006;104:c7–14.
- Cost of HRQOL from COOP/WONCA Charts. [cited 2014 Dec 9]. <http://qol.thoracic.org/sections/instruments/ae/pages/coop.html>
- Skevington SM. Measuring quality of life in Britain: introducing the WHOQOL-100. *J Psychosom Res* 1999;47:449–59.
- Skevington SM, Sartorius N, Amir M. Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL instruments. *Soc Psychiatry Psychiatr Epidemiol* 2004;39:1–8.
- Cost of WHOQOL. [cited 2014 Dec 9]. <http://www.bath.ac.uk/whoqol/questionnaires/info.shtml>
- Romney W, Weisbach C. Rehab Measures: Sickness Impact Profile. 2013. [cited 2014 Dec 9]. <http://www.rehabmeasures.org/Lists/RehabMeasures/PrintView.aspx?ID=955>
- Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
- Cost of SF-12. [cited 2014 Dec 9]. <https://www.optum.com/optum-outcomes/what-we-do/health-surveys/sf-12v2-health-survey.html>
- Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004–2011. *BMC Med Res Methodol* 2013;13:49.
- Antonovsky A. The structure and properties of the sense of coherence scale. *Soc Sci Med* 1993;36:725–33.
- Cost of SOC13 and SOC29. [cited 2014 Dec 9]. http://www.salutogenesis.hv.se/eng/SOC_questionnaire.19.html
- Bishop F, Yardley L. The development and initial validation of a new measure of lay definitions of health: the wellness beliefs scale. *Psychol Health* 2010;25:271–87.
- Eurich DT, Johnson JA, Reid KJ, et al. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health Qual Life Outcomes* 2006;4:89.
- Feeny D, Farris K, Côté I, et al. A cohort study found the RAND-12 and Health Utilities Index Mark 3 demonstrated construct validity in high-risk primary care patients. *J Clin Epidemiol* 2005;58:138–41.
- Johnson JA, Maddigan SL. Performance of the RAND-12 and SF-12 summary scores in type 2 diabetes. *Qual Life Res* 2004;13:449–56.
- Ware JE. SF-36® Health Survey Update. 2002. [cited 2014 Dec 9]. <http://www.sf-36.org/tools/SF36.shtml>
- Cost of SF-36. [cited 2014 Dec 9]. <https://www.optum.com/optum-outcomes/what-we-do/health-surveys/sf-36v2-health-survey.html>
- Pavot W, Diener E. Review of the satisfaction with life scale. *Psychol Assess* 1993;5:164–72.
- Cost of SWLS. [cited 2014 Dec 9]. <http://internal.psychology.illinois.edu/~ediener/SWLS.html>
- Aqtash S, Van Servellen G. Determinants of health-promoting lifestyle behaviors among Arab immigrants from the region of the Levant. *Res Nurs Health* 2013;36:466–77.
- Cost of HPLP-II. [cited 2014 Dec 9]. http://www.unmc.edu/nursing/docs/HPLPII_Background_and_Permission_for_Use.pdf
- Coons SJ. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000;17:13–35.
- Cost of EQ-5D (Euroqol). [cited 2014 Dec 9]. <http://www.euroqol.org/eq-5d-products/how-to-obtain-eq-5d.html>
- Cost of QWB. [cited 2014 Dec 9]. <https://hoap.ucsd.edu/qwb-info/QWB-Manual.pdf>
- Chang CY, MyersJE. Cultural adaptation of the wellness evaluation of lifestyle: an assessment challenge. *Meas Eval Couns Dev* 2003;35:239.
- Hattie JAMJESTJ. A factor structure of wellness: theory, assessment, analysis, and practice. *J Coun Dev* 2004;82:354–64.
- Cost of WEL. [cited 2014 Dec 9]. <http://www.mindgarden.com/products/wells.htm>
- Ventegodt S, Henneberg E, Merrick J, et al. Validation of two global and generic quality of life questionnaires for population screening: SCREENQOL and SEQOL. *Sci World J* 2003;3:412–21.
- Hoboken. *SCL-90-R*. The concise Corsini encyclopedia of psychology and behavioral science. John Wiley and Sons Inc, Hoboken, New Jersey. 2004.
- Schmitz N, Hartkamp N, Kiuse J, et al. The Symptom Check-List-90-R (SCL-90-R): a German validation study. *Qual Life Res* 2000;9:185–93.
- Vallejo MA, Jordán CM, Díaz MI, et al. Psychological assessment via the internet: a reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the symptoms check-list-90-revised (SCL-90-R). *J Med Internet Res* 2007;9:e2.
- Cost of SCL-90-R. [cited 2014 Dec 9]. <http://www.pearsonclinical.com/psychology/products/100000645/symptom-checklist-90-revised-scl90r.html>
- Hawthorne G. Measuring social isolation in older adults: development and initial validation of the friendship scale. *Soc Indicators Res* 2006;77:521–48.
- Cost of HUI3. [cited 2014 Dec 9]. <http://www.healthutilities.com/fees.htm>
- Kind P, Carr-Hill R. The Nottingham health profile: a useful tool for epidemiologists? *Soc Sci Med* 1987;25:905–10.
- Thorsen H, McKenna SP, Gottschalk L. The Danish version of the Nottingham Health Profile: its adaptation and reliability. *Scand J Prim Health Care* 1993;11:124–9.
- Cost of NHP. [cited 2014 Dec 9]. <http://www.rehabmeasures.org/Lists/RehabMeasures/DispForm.aspx?ID=931>
- Serdal KK, Derya Ö, Şehim K, et al. Psychometric properties of the Health Assessment Questionnaire Disability Index (HAQ-DI) and the Modified Health Assessment Questionnaire (MHAQ) in patients with knee osteoarthritis. *Arch Rheumatol* 2010;25:147–55.

- 45 Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). *Arthritis Care Res (Hoboken)* 2011;63(S11):S4–13.
- 46 PAR-QoL: Life Satisfaction Index-A (LSIA). [cited 2014 Dec 9]. <http://www.parqol.com/page.cfm?id=110>
- 47 Neugarten BL, Havighurst RJ, Tobin SS. The measurement of life satisfaction. *J Gerontol* 1961;16:134–43.
- 48 Browne J, O'Boyle C, McGee H, *et al.* Development of a direct weighting procedure for quality of life domains. *Qual Life Res* 1997;6:301–9.
- 49 Joyce C, O'Boyle C, McGee H. The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). 2003. [cited 2014 Dec 9]. https://www.google.com.au/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0CC0QFjAD&url=http%3A%2F%2Fwww.proqolid.org%2Fproqolid%2Fcontent%2Fdownload%2F4454%2F36412%2Ffile%2Finfoaboutseiqol.pdf&ei=CA6AVP2nJMWf8gWlwoDYCQ&usg=AFQjCNF-4Tma_osqSuQqraBjdYHjyBGJCQ&sig2=o9G7___kO08QEceQSFNW3gw&bvm=bv.80642063,d.dGc
- 50 O'Boyle C, McGee H, Hickey A, *et al.* The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). *Adm Man* 1993. [cited 2014 Dec 9]. <http://epubs.rcsi.ie/cgi/viewcontent.cgi?article=1041&context=psycholrep>
- 51 Wettergren L, Kettis-Lindblad Å, Sprangers M, *et al.* The use, feasibility and psychometric properties of an individualised quality-of-life instrument: a systematic review of the SEIQoL-DW. *Qual Life Res* 2009;18:737–46.