BMJ Open

Comparison between different instruments for measuring health-related quality of life in a population sample, the WHO MONICA Project, Gothenburg, Sweden: an observational, cross-sectional study

Emily Krantz, [1,2] Ulla Wide, [3] Penelope Trimpou, [2,4] Inger Bryman, [5,6] Kerstin Landin-Wilhelmsen [2,4]

For numbered affiliations see end of article.

Correspondence to
Dr Emily Krantz;
emily.amundson@vgregion.se

ABSTRACT

Objective The general aim was to meet the need for empirical comparative studies of health-related quality of life (HRQoL) assessment instruments, by evaluating and comparing the psychometric properties and results of three different, widely used, generic HRQoL instruments in a population sample. The specific aims were to evaluate the subscales of the different instruments that measure the same domain and to assess the association between the HRQoL measures and a single-item self-rated health scale.

Design An observational cross-sectional study.

Setting A population-based sample from Gothenburg, Sweden, was studied in 2008 in the WHO MONItoring of trends and determinants for CArdiovascular disease.

Participants A total of 414 subjects were included, 77% women, age range 39–78 years.

Interventions The Nottingham Health Profile (NHP), the Short Form-36 questionnaire (SF-36), the Psychological General Well-Being Index (PGWB) and a self-rated health scale were used.

Outcome measures Scores were analysed for their psychometric properties, internal consistency (Cronbach's α), construct validity (Spearman's rank correlations and R^2 coefficients) and discriminative ability for the presence of self-rated ill-health.

Results PGWB and SF-36 had higher Cronbach's α scores than NHP. All correlations calculated between the subscales that were conceptually similar were significant (p<0.01). All subscales could differentiate the presence of self-rated ill-health according to the self-rated health scale (p<0.001). The self-rated health scale correlated strongly with all of the three HRQoL instruments used.

Conclusions There was a high concordance between the instruments within each domain that was conceptually similar. All three HRQoL instruments (PGWB, SF-36 and NHP) could discriminate the presence of self-rated ill-health. The simple and quick self-rated health scale correlated strongly with the more time-consuming PGWB, SF-36 and NHP. The result supports the existence of a strong association between the self-rated health scale and HRQoL in the general population.

Strengths and limitations of this study

► Health-related quality of life (HRQoL) measurements are frequently asked for in clinical trials and controlled studies. Very few comparative, methodological studies of HRQoL instruments have been published on population samples. This study reports the results of three different HRQoL instruments from the general population of over 400 Swedish subjects with a largely complete data set.

► All subjects completed the questionnaires in the same order minimizing the risk for systematic error.

► The definition of ill-health was self-rated using a quick and simple single-item self-rated health scale (0–100), which may affect conclusions about the discriminant validity of the instruments.

► The cross-sectional design makes it impossible to report on the responsiveness of the instruments, which is an important criterion when evaluating an HRQoL instrument. Content validity, structural validity and measurement error were not evaluated either.

► The study population is comprised of middle-aged, elderly, predominantly female subjects, which affects the generalizability of the sample.

INTRODUCTION

Health-related quality of life (HRQoL) is an important variable in clinical practice and in the medical literature with significant consequences for patients and society. As the general population ages and treatments become more advanced, widespread and expensive, interest has grown in evaluating medical treatments using patient-reported outcome measures, such as self-assessed HRQoL, as key variables.[1 2] HRQoL has become an integral part of medical clinical research in all disciplines, and is even seen as

Table 1 Comparison of the content of the Nottingham Health Profile (NHP), Short Form-36 (SF-36), Psychological General Well-Being Index (PGWB) and the self-rated health scale to identify domains that are conceptually similar

Domain name	NHP	SF-36	PGWB	Self-rated health scale
Social functioning	Social isolation	Social functioning	–	–
Pain	Pain	Bodily pain	–	–
Physical functioning	Physical mobility	Physical functioning	–	–
Mental health	Emotional reactions	Mental health	Anxiety and Depressed mood	–
Vitality	Energy	Vitality	Vitality	
General health	–	General health	General health	Self-rated health
Summary scores	–	Physical component summary and Mental component summary	PGWB total score	Self-rated health

a hard end-point, alongside survival.[3 4] However, a major challenge has been to find widely accepted definitions of HRQoL.[5]

HRQoL is, by nature, subjective and a multidimensional approach must be taken to encompass physical and occupational function, psychological state, social interaction and somatic sensation caused by an illness and its consequent therapy on a patient.[6] HRQoL instruments are generally used to quantify health into health dimensions, or domains, such as mobility, ability to perform certain activities, emotional state, sensory function, cognition, social function and freedom from pain.[1]

There is a growing number of HRQoL measurement instruments available to researchers, and their sophistication, variety and scope is increasing. Since comparisons between clinical groups and population samples are common, it is important that the HRQoL instruments used are reliable and valid in the population. However, few studies apply different instruments and compare the results, and even fewer do so in general population samples. A meta-analysis planned by Lorente *et al* aims to evaluate HRQoL instruments indicating the need for such comparisons.[7]

Studies done in Dutch population samples in 1996[8] and 1997[9] and in a Brazilian population sample in 2011[10] are examples of studies that have applied different HRQoL instruments. All aimed to compare the reliability of scores, to assess the discriminative ability of potential outcome measures applied in a general population sample and to assess the extent of agreement between the different instruments. The authors concluded that it is important to define one's research question and underlined the need for careful consideration when choosing among HRQoL instruments. However, this is difficult when head-to-head analyses of different instruments with overlapping purposes are so rare. The HRQoL instruments compared in this study are the Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB) and the

Medical outcomes study Short Form-36 (SF-36). All of the instruments reflect the HRQoL domains outlined above.

The general aim was to meet the need for empirical comparative studies of HRQoL assessment instruments, by evaluating and comparing the psychometric properties and results of three different, widely used, generic HRQoL instruments in a population sample. A specific aim was to evaluate the subscales of the different instruments that measure the same HRQoL domain. The hypothesis was that there would be a high concordance between similar subscales in the different instruments. Another specific aim was to assess the association between the HRQoL instruments and an easily administered single-item self-rated health scale. The hypothesis was that the self-rated health scale is strongly associated with all the domains of HRQoL.

METHODS
Study setting
This is an observational cross-sectional study of a population-based sample, n=414, from Gothenburg, Sweden.

Sample selection process
In 1995, 2592 individuals (age 25–64, 50% women) were recruited from the Gothenburg city census, which is kept up to date within a maximum of 14 days. This was the third population screening by the WHO MONICA-GOT (WHO MONItoring of trends and determinants for CArdiovascular disease GOThenburg) in which 1618 individuals participated.[11] The non-attenders in 1995 could not participate due to travel, living abroad, unwillingness to attend or inability to attend due to the illness of a relative. The subjects were examined at a medical clinic. A randomly selected subset of these subjects (every fourth subject, and all of the women aged 45–64 years, in total 662) underwent extra testing and they were invited for re-evaluation and assessment of HRQoL in 2008.[12]

Table 2 Descriptive statistics and features of the Psychological General Well-Being Index (PGWB), Nottingham Health Profile (NHP), Short Form-36 (SF-36) and the self-rated health scale

All subjects

	Mean (SD)	Median	Floor (%)	Ceiling (%)	Cronbach's α
PGWB (n=409)					
Anxiety	24.1 (5.1)	26	0	34	0.90
Depressed mood	15.9 (2.7)	17	0.2	42	0.88
Positive well-being	16.4 (3.8)	17	0	1	0.85
Self-control	15.5 (2.6)	16	0	18	0.76
General health	14.3 (3.2)	15	0.2	16	0.76
Vitality	17.3 (4.2)	18	0.5	3	0.87
Total score	103.6 (18.6)	109	0	1	–
NHP* (n=409)					
Emotional reaction	90.0 (18.8)	100	0.7	63	0.83
Sleep	78.7 (27.2)	89	3	44	0.67
Energy	83.3 (30.6)	100	7	70	0.77
Pain	85.8 (25.2)	100	2	60	0.87
Physical mobility	91.0 (17.0)	100	0.5	66	0.80
Social isolation	94.3 (15.3)	100	0.2	84	0.66
SF-36 (n=407)					
Physical functioning	78.0 (23.7)	85	0.7	23	0.91
Role physical	74.5 (37.3)	100	14	60	0.88
Bodily pain	69.5 (25.7)	72	1	27	0.85
General health	69.3 (23.4)	72	0.7	8	0.83
Vitality	66.1 (23.6)	70	1	5	0.85
Social functioning	85.6 (22.7)	100	1	59	0.84
Role emotional	78.8 (35.3)	100	11	69	0.84
Mental health	77.6 (19.9)	84	0.2	10	0.86
Physical component summary	48.7 (10.3)	51	–	–	–
Mental component summary	52.0 (11.2)	56	–	–	–
Self-rated health scale (n=403)	75.7 (20.4)	80	0.2	5	–

Floor/ceiling effects are not relevant for SF-36 mental and physical component summaries because these scores are calculated using US-norm values from 1998.
Cronbach's α coefficient is not relevant for PGWB total score, SF-36 mental or physical component summaries and the self-rated health scale and is, therefore, not shown.
*NHP scores are reversed for consistency with the other instruments.

Of these subjects, 495 responded, 97 were deceased, 13 could not be traced and 57 did not reply. Sixty-four declined consent to participate and 17 did not come to the clinic. In total, 414 subjects completed the HRQoL questionnaires. Two subjects were excluded because of incomplete data, leaving 412 subjects who were included in the analysis (62% participation rate, 77% women, age range 39–78 years).

Procedure

The subjects completed the questionnaires while visiting the Sahlgrenska University Hospital, Gothenburg, for medical examinations. After blood sampling, all subjects received breakfast during which the questionnaires were administered in the following order: NHP, PGWB, SF-36 and a single-item self-rated health scale. A single operator performed the measurements and administrations on all subjects. No personal guidance was given except for the instructions.

HRQoL instruments

Nottingham Health Profile

NHP measures aspects of subjective health using a two-part questionnaire.[13] In this study, the NHP part I was used. Part I is comprised of 38 statements covering six dimensions concerning distress or limitations of activity: physical mobility, pain, sleep, energy, social isolation and emotional reactions. The response format is yes or no, dimension
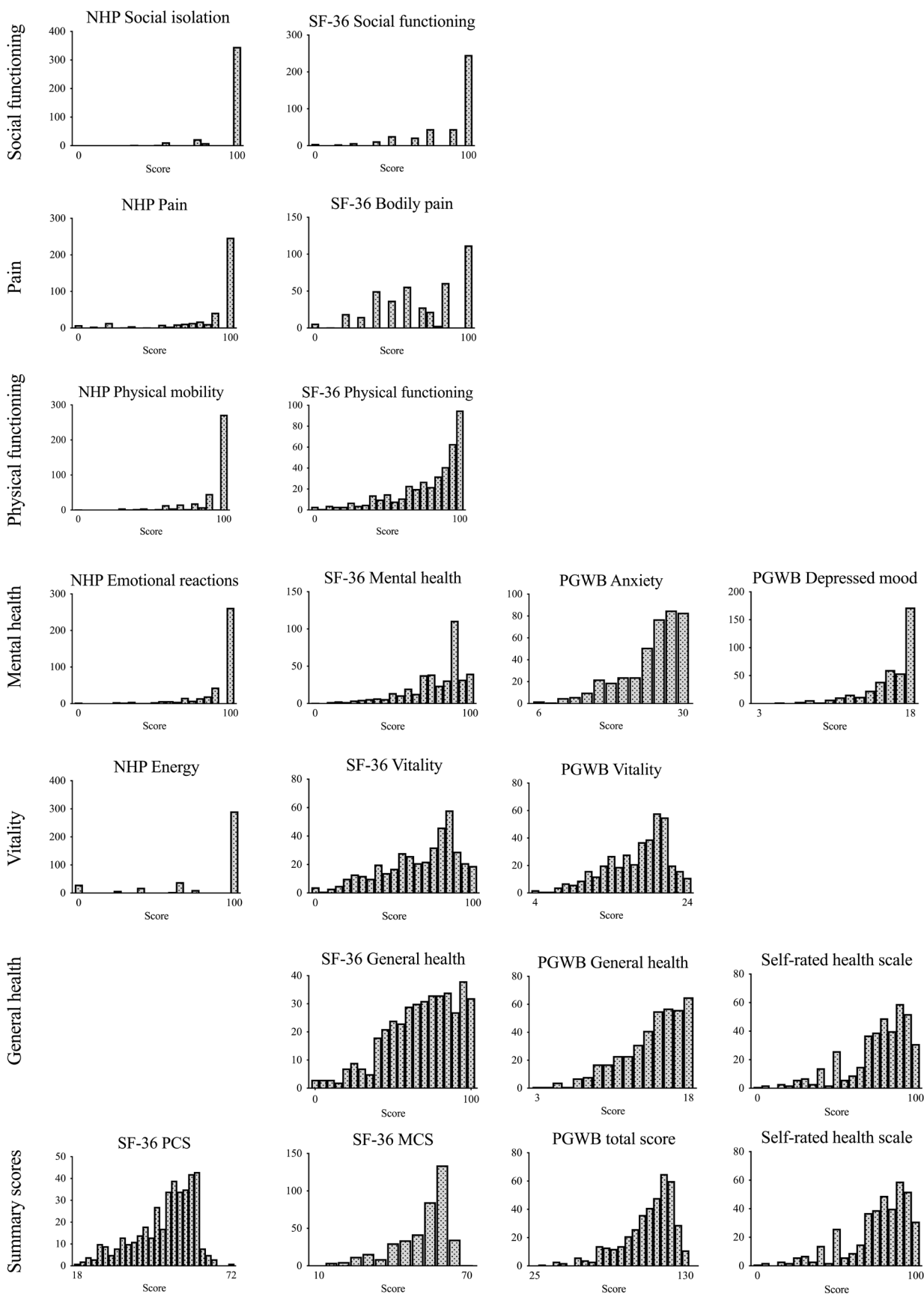
Figure 1 Frequency distributions (histograms) of similar subscales in the Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB), Short Form-36 (SF-36) and the self-rated health scale. Each row representing a domain (from top to bottom): social functioning, pain, physical functioning, mental health, vitality, general health and summary scores.

Table 3 Spearman's rank correlations between the similar subscales (shown in table 1) in the Psychological General Well-Being Index (PGWB), Nottingham Health Profile (NHP), Short Form-36 (SF-36) and the self-rated health scale (SRHS)

	PGWB					NHP						SF-36							
	Anxiety	Depressed mood	General health	Vitality	Total score	Emotional reactions	Energy	Pain	Physical mobility	Social isolation	SRHS	Physical functioning	Bodily pain	General health	Vitality	Social functioning	Mental health	PCS*	MCS†
PGWB																			
Anxiety	1																		
Depressed mood	**0.70**	1																	
General health	0.45	0.49	1																
Vitality	0.68	0.66	0.65	1															
Total score	0.84	0.82	0.70	0.90	1														
NHP‡																			
Emotional reactions	**0.67**	**0.70**	0.48	0.66	0.75	1													
Energy	0.42	0.51	0.53	**0.59**	0.57	0.59	1												
Pain	0.31	0.37	0.63	0.42	0.47	0.36	0.46	1											
Physical mobility	0.28	0.35	0.55	0.41	0.43	0.36	0.51	0.66	1										
Social isolation	0.35	0.43	0.32	0.35	0.42	0.48	0.45	0.27	0.29	1									
SRHS	0.49	0.54	**0.66**	0.63	**0.68**	0.53	0.55	0.46	0.48	0.36	1								
SF-36																			
Physical functioning	0.27	0.39	0.62	0.41	0.46	0.38	0.52	0.67	**0.72**	0.28	0.58	1							
Bodily pain	0.33	0.37	0.75	0.48	0.52	0.36	0.41	**0.67**	0.56	0.24	0.55	0.67	1						
General health	0.43	0.49	**0.66**	0.57	0.62	0.50	0.57	0.55	0.54	0.35	**0.77**	0.65	0.56	1					
Vitality	0.64	0.62	0.66	**0.84**	0.83	0.63	**0.61**	0.47	0.46	0.37	0.70	0.49	0.53	0.66	1				
Social functioning	0.55	0.59	0.56	0.59	0.67	0.55	0.49	0.37	0.33	**0.37**	0.55	0.40	0.40	0.50	0.59	1			
Mental health	**0.72**	**0.74**	0.52	0.70	0.82	**0.70**	0.48	0.36	0.30	0.41	0.57	0.35	0.40	0.54	0.76	0.59	1		
PCS*	0.23	0.30	0.73	0.43	**0.44**	0.29	0.48	0.68	0.68	0.22	**0.61**	**0.86**	0.81	0.73	0.51	0.37	0.26	1	
MCS†	0.68	0.69	0.40	0.67	**0.76**	0.65	0.45	0.23	0.18	0.40	**0.50**	0.16	0.24	0.45	0.73	0.66	0.89	0.08§	1

Coefficients in similar subscales are shown in bold.
*Physical component summary.
†Mental component summary.
‡NHP scores are reversed for consistency with the other instruments.
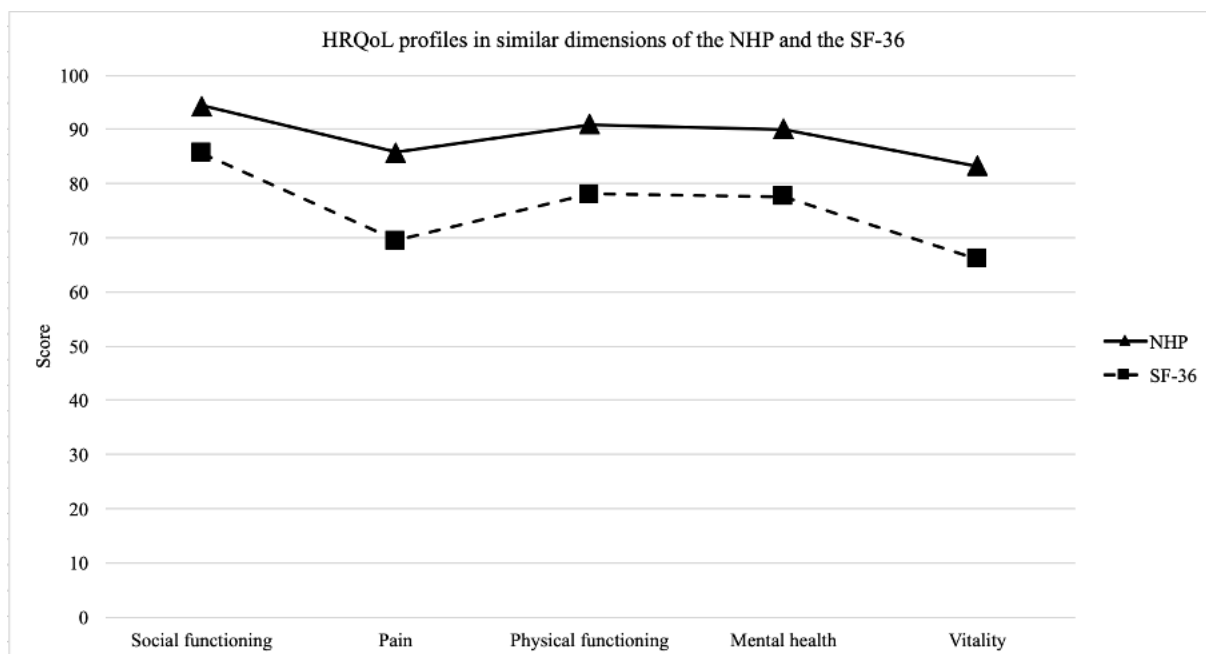§p>0.05, NOT significant. All other correlations are significant at p<0.01.

Figure 2 Distributions of scores in comparable dimensions in the Nottingham Health Profile (NHP) and Short Form-36 (SF-36). NHP scores are reversed for consistency with the other instruments.

scores range from 0 to 100 and each statement is weighted according to the level of severity. The higher the score, the greater the limitations/distress, that is, the lower HRQoL. The NHP was developed in the 1980s but is still widely used, especially in Europe. It is useful because of its breadth and simplicity and is a suitable instrument for use in clinical practice and in populations where there are likely to be people with disabilities.[14]

Psychological General Well-Being Index
The PGWB was designed to measure personal affective or emotional states reflecting a sense of well-being or
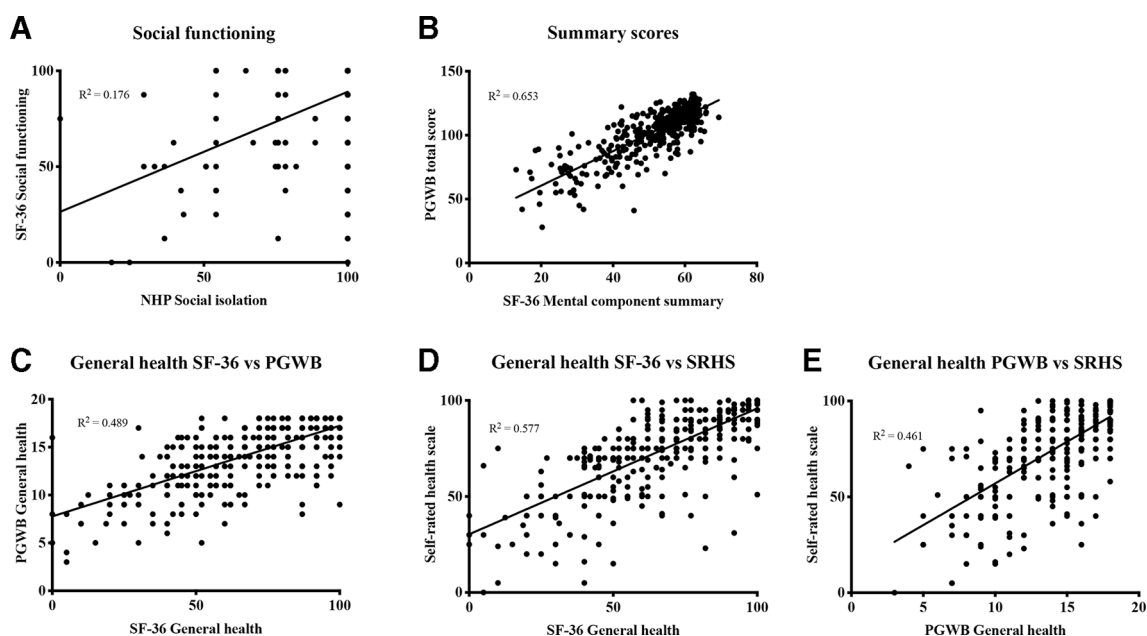
Figure 3 Correlation scatterplots between similar subscales in the Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB), Short Form-36 (SF-36) and the self-rated health scale (SRHS). (A) Social functioning SF-36 versus NHP. (B) Summary score SF-36 MCS versus PGWB. (C) General health SF-36 versus PGWB. (D) General health SF-36 versus self-rated health scale. (E) General health PGWB versus self-rated health scale. R^2=coefficient of determination (goodness of fit). MCS, mental component summary; NHP scores are reversed for consistency with the other instruments.

Table 4 Discriminative ability of the Psychological General Well-Being Index (PGWB), Nottingham Health Profile (NHP), and Short Form-36 (SF-36) to identify ill-health. Mean values are given for all subscales when split according to the self-rated health scale median score=80

Self-rated health—split at median	0–79	80–100	Effect size d*
PGWB			
Anxiety	21.7	26.0	0.9
Depressed mood	14.5	17.1	1.1
Positive well-being	14.2	18.1	1.2
Self-control	14.3	16.5	1.6
General health	12.1	16.0	1.5
Vitality	14.8	19.2	1.4
PGWB total score	91.3	112.9	1.4
NHP†			
Emotional reaction	80.4	97.1	1.0
Sleep	67.2	87.2	0.8
Energy	65.7	96.0	1.1
Pain	72.1	96.0	1.1
Physical mobility	82.5	97.2	1.0
Social isolation	89.0	98.3	0.6
SF-36			
Physical functioning	63.4	88.8	1.3
Role physical	50.3	92.4	1.4
Bodily pain	53.3	81.7	1.3
General health	51.0	82.9	1.9
Vitality	49.6	78.3	1.5
Social functioning	73.5	94.6	1.1
Role emotional	60.4	92.6	1.0
Mental health	65.8	86.2	1.2
Physical component score	41.7	53.9	1.4
Mental component score	46.0	56.4	1.1

p<0.001 for comparisons in all subscales, calculated using Mann-Whitney U test.
*Cohen's d test for standardised mean effect size, d>0.5 considered clinically significant.
†NHP scores are reversed for consistency with the other instruments.

distress intended for use in community surveys.[15] The PGWB includes 22 items, with a six-grade Likert style response format where a high score represents a better HRQoL. The scores are summarised into an overall well-being score (PGWB total score, range 22–132), and is also divided into six subscales: anxiety (range 5–30), depressed mood (range 3–18), positive well-being (range 4–24), self-control (range 3–18), general health (range 3–18) and vitality (range 4–24). The PGWB has been used in clinical trials and has performed well in both population-based and mental health samples.[16]

Medical outcomes study SF-36 questionnaire

The SF-36 is a multipurpose health survey comprised of 36 items where a high score represents a better HRQoL.[17] It yields an eight-scale profile of functional health and well-being: physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional and mental health (range for all 0–100). It also generates psychometrically based physical and mental health summary measures: a mental component summary and a physical component summary. The mental component summary is comprised of the subscales for vitality, social functioning, role emotional and mental health, whereas the physical component summary is comprised of the subscales for physical functioning, role physical, bodily pain and general health. The SF-36 has been proven useful in surveys of general and specific populations, comparing the relative burden of diseases, and in differentiating the health benefits produced by a wide range of different treatments.[18]

Self-rated health scale

Self-rated health was measured with a single question. Subjects were asked to rate their current health status between 0 and 100 on a linear analogue self-assessment scale; 0 being the worst conceivable level and 100 being the best conceivable level. The item is identical

to question number six published in the 1990 edition of the EuroQoL-5 Dimension questionnaire, EQ-5D.[19] Such single-item health indicators have consistently been shown to be strong correlates of objective health and even as predictors of mortality.[20–22]

Background variables

Age in whole years and sex were determined using the Swedish personal identity number on the day of the visit. Information about education level was recorded in whole years from the first grade, according to the subject.

Statistical methods

Descriptive statistics for each of the instrument's subscales including mean, median, SD and percentage of subjects with lowest (floor effect) and highest (ceiling effect) possible scores were calculated. The non-parametric Mann-Whitney U test was used to comapare continuous vairables, since the results were not normally distributed. The standardized mean effect size was calculated using Cohen's d test (mean difference divided the pooled variance), d>0.25 was considered educationally significant and d>0.5 was considered clinically significant.[23] Internal consistency was examined using Cronbach's α, α>0.70 was considered acceptable. Correlation analyses between the instruments were focused on comparing the conceptually similar dimensions between the instruments used. Spearman's rho correlations (r_s) were used to analyse discriminant validity since the results were not normally distributed. Correlation coefficients were considered weak if r_s<0.30, moderate if r_s=0.30–0.49 and strong if r_s≥0.50. Regression analysis using the R^2 coefficient of determination was also calculated for certain subscale comparisons. The presence of self-rated ill-health was defined using the self-rated health scale score split at the median. All scores below the median value were categorized as self-rated ill-health.

All statistical analyses were calculated using Statistical Package for the Social Sciences (V.24) software or Microsoft Excel. A p value of <0.01 was chosen to reduce the risk of type II error. SF-36 scores were calculated using scoring software obtained from Optum (license number QM03712) and, mental and physical component scores were calculated using 1998 US norms. NHP scores were reversed for consistency with the other instruments to facilitate comparisons.

Missing values were imputed in NHP questionnaire if less than 80% of the values were missing in a given subscale. In these 20 instances, the median value was calculated and imputed. Imputing was considered unnecessary when analysing the PGWB and the self-rated health scale because the sample size was large and missing answers were not common.

In order to compare the results between the instruments NHP, PGWB, SF-36 and the self-rated health scale, the authors identified six domains that were conceptually similar: social functioning, pain, physical functioning, mental health, vitality and general health, and

the summary scores. This categorisation was made based on the content in the items themselves and supported by previously published studies using these instruments[10 24–27] (table 1).

The STrengthening the Reporting of OBservational studies in Epidemiology cross-sectional checklist was used when writing the report.[28]

Participant involvement

No subjects were involved in any stage of development, implementation or interpretation of this study. There are no plans to disseminate the results of this study to the study subjects.

RESULTS
Characteristics of the subjects

The mean age of the subjects who were included in the analysis (n=412) was 62.8 years, range 39–78. Seventy-seven per cent were women with a mean age of 63.7 years, the men had a mean age of 59.6 years (p<0.001). The average number of school years was 12, no significant difference was found between men and women (data not shown). Most of the subjects (>90%) had been employed but were retired at the time of this investigation.

Questionnaire scores and distribution

Three subjects (1%) had incomplete or largely incomplete NHP and PGWB questionnaires, five subjects (1.2%) had incomplete or largely incomplete SF-36 questionnaires and nine subjects (2%) did not complete the self-rated health scale.

Descriptive statistics for each of the HRQoL instruments are presented for the whole group in table 2. Men and women scored similarly in all the NHP subscales and the self-rated health scale. There were statistically significant differences between the sexes in some of the PGWB and SF-36 subscales, but further analysis to determine the effect size showed none of these differences to be of clinical significance (Cohen's d range 0.2–0.4) (data not shown).

The distribution of the results was skewed for all of the instruments (figure 1). The ceiling effect was most prominent in the NHP, in which 43%–84% of the respondents scored at the ceiling in the different subscales. The highest proportion of respondents scoring at the ceiling in the NHP subscales was in the subscales social isolation (84%), energy (70%) and physical mobility (66%). The highest ceiling effects in the SF-36 were seen in the subscales role emotional (69%), role physical (60%) and social functioning (59%). The highest proportion of ceiling scores in PGWB was seen in the subscale depressed mood (42%). The self-rated health scale was the least skewed of all the instruments used and only 5.3% reported the highest possible score of 100 (table 2).

Reliability

Internal consistency coefficients for all instruments are shown in table 2. The NHP yielded lower internal consistency estimates than the other two instruments (NHP mean α=0.77, range 0.66–0.87; PGWB mean α=0.85, range 0.76–0.90; SF-36 mean α=0.86, range 0.83–0.91). Two of the subscales in the NHP fell below the standard recommended α>0.70 for group comparisons (social isolation and sleep). All of the eight SF-36 subscales and four of the six subscales in the PGWB had α-coefficients >0.80.

Convergent validity

Correlations between the relevant subscales (table 1) of the PGWB, SF-36, NHP and the self-rated health scale are shown in table 3. Correlation coefficients for the conceptually similar subscales are shown in bold. All interinstrument correlations were significant at p<0.01 in the hypothesised direction.

Correlations between similar domains within NHP and SF-36

The results found in the comparable dimensions of the SF-36 and NHP are shown in figure 2. There were positive correlations between all the similar subscales of the SF-36 and NHP (physical functioning, pain, vitality, social functioning and mental health, all p<0.01) (table 3).

Correlations between subscales measuring similar domains within NHP, SF-36 and PGWB

The correlations were positive and strong in the subscales between the PGWB, SF-36 and NHP, respectively, in the domains they had in common (mental health, p<0.01; vitality, p<0.01). The PGWB subscales were more strongly associated with the SF-36 subscales than with the NHP subscales. Furthermore, the associations between the PGWB and the SF-36 were stronger than the associations between SF-36 and NHP within these domains. The PGWB total score was associated with both the SF-36 summary scores but the association was weaker with the physical component summary than with the mental component summary.

Correlations between the self-rated health scale and the HRQoL instruments

Correlations between the self-rated health scale and the general health subscales in the PGWB and the SF-36 were strong. The associations between the self-rated health scale and the PGWB total score and the SF-36 physical component summary and mental component summary were also strong. It is notable that there were no weak correlations between the self-rated health scale and *any* of the other instruments' subscales (table 3).

Scatterplot diagrams were used to examine and visualise the relationships between some of the subscales (figure 3). The social functioning domain (figure 3A) showed an R^2 coefficient of 0.18 for the NHP versus the SF-36, meaning that only approx. 18% of the variation in social functioning measured with the NHP is described by the change in the same dimension measured with the SF-36. The correlation between the PGWB total score and

the SF-36 mental component summary was strong and the linear relationship is the highest of all the comparisons tested, with an R^2=0.65 (figure 3B). The general health domain also showed a strong correlation between all three instruments with the highest R^2 coefficient between the self-rated health scale and the SF-36 general health (R^2=0.58) (figure 3C–E).

Discriminative ability

To compare the ability of the PGWB, the NHP and the SF-36 instruments to discriminate subjects on the basis of health, the presence of ill-health was defined as self-rated health scale <80 (median score). All of the subscales could significantly differentiate the presence of self-perceived ill-health (p<0.001) and the effect sizes for all subscales were above the threshold to be considered clinically significant (Cohen's d>0.5) (table 4).

DISCUSSION

The general aim of the study was to examine and compare the psychometric properties of three generic HRQoL instruments—the NHP, the SF-36 and the PGWB—and their association to the self-rated health scale when used in a general population sample. The instruments showed strong reliability and discriminative ability, and the subscales measuring the same HRQoL domain showed strong associations (mainly r_s>0.60), except in the social functioning domain. The distributions were skewed with considerable ceiling effects, which is to be expected when measuring HRQoL in a general population sample. All instruments differentiated between individuals with poor and good health.

It is widely accepted from psychometric literature that an HRQoL-measurement's quality can be judged on the reliability, stability, prominence of ceiling/floor effects and validity.[29] Stability was not tested here because of the cross-sectional nature of this study, but by the other criteria mentioned, the SF-36 and the PGWB performed equally well and both performed slightly better than the NHP. The PGWB had equivalent internal consistency to the SF-36, and had the least prominent ceiling and floor effect of the HRQoL instruments used.

Strong correlations were found between the PGWB and the SF-36 in the mental health, general health and vitality domains as well as between the PGWB total scores and SF-36 mental component summary. A study on patients with asthma also found a high correlation between the SF-36 mental component summary and the PGWB total score, and concluded that administering the PGWB together with the SF-36 would be redundant.[30] Another study in patients with amyotrophic lateral sclerosis that focused solely on the mental health subscales in these two instruments found that internal consistency was equivalent and that all the PGWB subscales correlated strongly with the SF-36 mental health subscale.[31] In the present study, a more nuanced approach was taken by comparing several similar subscales and not singling out mental

health. These results support earlier recommendations to choose one or the other, particularly when the goal is to assess mental health, general health or vitality in a population sample in which the majority of the subjects do not have a chronic disease.

The PGWB had the same ability to discriminate the presence of self-rated ill-health as the SF-36 and the NHP. However, the PGWB should perhaps not stand alone if the aim is to assess HRQoL in a population since it does not meet the customary criteria for an HRQoL instrument.[32] It does, on the other hand, contain aspects of positive well-being that the others may miss.[16] These results are important, first because this is the only study, to the best of our knowledge, that compares the PGWB to the SF-36 applied in a general population sample, and second, because there is a lack of validity studies for the PGWB.[14]

In the present study, the SF-36 had a higher internal consistency, less prominent floor/ceiling effects and less skewed results than the NHP. These results support earlier findings that the SF-36 performs better than the NHP in population samples.[8 9] The congruity between the two instruments was weakest in the social isolation domain. The SF-36 social functioning subscale was more strongly associated with the NHP subscales for emotional reactions and energy—much like previous findings.[26 27] The items in the social functioning domain differ considerably in their content, which may explain this result.[33] The SF-36 includes two questions on how/if physical and/or mental problems affect social interactions. The NHP includes five items on loneliness, social interactions, close friends and a feeling of being a burden to others, but without the specific connection to physical or mental symptoms. Notably, for the four remaining common domains, covering both mental and physical aspects of HRQoL, each pair of NHP and SF-36 scales were strongly correlated. Both instruments had the same ability to discriminate the presence of self-rated ill-health.

The only previous population-based comparison, to the best of our knowledge, between the SF-36 and the NHP was performed by Faria *et al* in community-dwelling subjects in Brazil with a mean age of 70 years.[10] The results were mainly similar regarding internal consistency and convergent validity. Faria *et al* concluded that the SF-36 may be slightly favourable for use in a group of community-dwelling elders because of the prominent ceiling effects seen in NHP. Like Prieto *et al*, who studied patients with lung disease, we only found small differences between the instruments. It is questionable whether the small differences are clinically relevant even when the instruments are applied in a population sample.[27]

In this study, the self-rated health scale correlated, not only with similar subscales in the general health domain, but with *all* the other instruments' subscales. The correlations between the self-rated health scale and the NHP were the least pronounced of all the comparisons made, with moderate correlations for the subscales measuring the domains of sleep, pain, physical mobility and social isolation. As expected, this population sample did not report a high level of problems in these domains using the NHP. This makes it reasonable to conclude that these domains, when measured with the NHP, were not a major cause of distress for the subjects, and did not strongly affect how they rated their health with the self-rated health scale.

The self-rated health scale could be considered a measure of overall HRQoL even in general population samples when the need for quick and easy administration is pertinent.[14 34 35] However, a single-item self-rated health measurement cannot be seen as a substitute for multi-item questionnaires when more specific information about specific domains, such as mental functioning, sleep and pain, for example, are required.

Strengths and limitations

Very few comparative studies of HRQoL measurements have been published on population samples. This study reports the results from more than 400 subjects with a largely complete data set collected in 2008. However, the inclusion of middle-aged, mainly retired, predominantly female subjects may have led to selection bias and also affects the generalizability of the sample even if the follow-up rates were high. The conclusions about the discriminant validity of the instruments must also be drawn with care since the definition of ill-health was self-rated using the self-rated health scale. Another limitation is the cross-sectional design that makes it impossible to report on the responsiveness of the instruments. Content validity, structural validity or measurement error were not evaluated either, and are all important criteria when evaluating HRQoL instruments.[36] The order in which the instruments were administered could have resulted in a 'context effect-bias'. However, all subjects completed the questionnaires in the same order minimising the risk for systematic error.[20]

CONCLUSIONS

There was a high concordance between the instruments for evaluating HRQoL within each domain that was conceptually similar, except in the social functioning domain. The PGWB performed as well as the SF-36 and better than the NHP regarding internal consistency. All three instruments could discriminate the presence of self-rated ill or good health. The self-rated health scale score correlated significantly with all the other instruments' subscales. The results support the hypothesis of a strong association between self-rated health and HRQoL, and the single-item self-rated health scale should be considered when time and resource efficiency are required.

Author affiliations
[1]Clinic of Internal Medicine, Södra Älvsborgs Hospital, Borås, Sweden
[2]Department of Internal Medicine, Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[3]Department of Behavioral and Community Dentistry, Institute of Odontology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[4]Section of Endocrinology, Sahlgrenska University Hospital, Gothenburg, Sweden

[5]Department of Obstetrics and Gynecology, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
[6]Department of Reproductive Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden

REFERENCES

1. Parrish RG. Measuring population health outcomes. *Prev Chronic Dis* 2010;7:A71.
2. Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
3. Bottomley A, Pe M, Sloan J, *et al*. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clin Trials* 2018;15:624–30.
4. Kovic B, Jin X, Kennedy SA, *et al*. Evaluating progression-free survival as a surrogate outcome for health-related quality of life in oncology: a systematic review and quantitative analysis. *JAMA Intern Med* 2018;178:1586–96.
5. Dodge R, Daly A, Huyton J, *et al*. The challenge of defining wellbeing. *International Journal of Wellbeing* 2012;2:222–35.
6. Schipper H, Clinch JJ, Olweny CLM. Quality of Life Studies: Definitions and Conceptual Issues. Spilker B, ed. *Quality of life and pharmacoeconomics in clinical trials*. 2nd ed.. Lippincott-Raven: Philadelphia, Pennsylvania, USA, 1996:11–23.
7. Lorente S, Vives J, Viladrich C, *et al*. Tools to assess the measurement properties of quality of life instruments: a meta-review protocol. *BMJ Open* 2018;8:e022829.
8. VanderZee KI, Sanderman R, Heyink J. A comparison of two multidimensional measures of health status: the Nottingham health profile and the RAND 36-Item health survey 1.0. *Qual Life Res* 1996;5:165–74.
9. Essink-Bot ML, Krabbe PF, Bonsel GJ, *et al*. An empirical comparison of four generic health status measures. The Nottingham health profile, the medical outcomes study 36-item short-form health survey, the COOP/WONCA charts, and the EuroQol instrument. *Med Care* 1997;35:522–37.
10. Faria CD, Teixeira-Salmela LF, Nascimento VB, *et al*. Comparisons between the Nottingham health profile and the Short Form-36 for assessing the quality of life of community-dwelling elderly. *Rev Bras Fisioter* 2011;15:399–405.
11. Wilhelmsen L, Johansson S, Rosengren A, *et al*. Risk factors for cardiovascular disease during the period 1985-1995 in Göteborg, Sweden. The GOT-MONICA Project. *J Intern Med* 1997;242:199–211.
12. Trimpou P, Lindahl A, Lindstedt G, *et al*. Secular trends in sex hormones and fractures in men and women. *Eur J Endocrinol* 2012;166:887–95.
13. Wiklund I. The Nottingham Health Profile-a measure of health-related quality of life. *Scand J Prim Health Care Suppl* 1990;1:15–18.
14. McDowell I. *Measuring Health: a guide to rating scales and questionnaires*. 3rd ed. New York: Oxford University Press, 2006.
15. Dupuy H. The Psychological General Well-Being (PGWB) Index. In: Wenger N, Mattson M, Furberg C, eds. *Assessment of quality of life in clinical trials of cardiovascular therapies*. New York: Le Jacq Publishing Inc, 1984:170–83.
16. Naughton MJ, Wiklund I. A critical review of dimension-specific measures of health-related quality of life in cross-cultural research. *Qual Life Res* 1993;2:397–432.
17. Sullivan M, Karlsson J, Ware JE. The Swedish SF-36 Health Survey-I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Soc Sci Med* 1995;41:1349–58.
18. Persson LO, Karlsson J, Bengtsson C, *et al*. The Swedish SF-36 Health Survey II. Evaluation of clinical validity: results from population studies of elderly and women in Gothenborg. *J Clin Epidemiol* 1998;51:1095–103.
19. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
20. Lee S. Self-Rated Health in Health Surveys. Johnson TP, *Handbook of health survey methods*. 1st edn. Hoboken, New Jersey: John Wiley and Sons, Inc, 2015.
21. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21–37.
22. Bernhard J, Sullivan M, Hürny C, *et al*. Clinical relevance of single item quality of life indicators in cancer clinical trials. *Br J Cancer* 2001;84:1156–65.
23. Wolf FM. *Meta-analysis: quantitative methods for research synthesis*. Beverly Hills: Sage, 1986.
24. Zengin N, Ören B, Gül A, *et al*. Assessment of quality of life in haemodialysis patients: a comparison of the Nottingham Health Profile and the Short Form 36. *Int J Nurs Pract* 2014;20:115–25.
25. Klevsgård R, Fröberg BL, Risberg B, *et al*. Nottingham health profile and short-form 36 health survey questionnaires in patients with chronic lower limb ischemia: before and after revascularization. *J Vasc Surg* 2002;36:310–7.
26. Meyer-Rosberg K, Burckhardt CS, Huizar K, *et al*. A comparison of the SF-36 and nottingham health profile in patients with chronic neuropathic pain. *Eur J Pain* 2001;5:391–403.
27. Prieto L, Alonso J, Ferrer M, *et al*. Are results of the SF-36 health survey and the Nottingham Health Profile similar? A comparison in COPD patients. Quality of Life in COPD Study Group. *J Clin Epidemiol* 1997;50:463–73.
28. von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
29. Polit D. *Nursing research: principles and methods*. 7th edn. Philadelphia: Lippincott Williams and Wilkins, 2004.
30. van der Molen T, Postma DS, Schreurs AJ, *et al*. Discriminative aspects of two generic and two asthma-specific instruments: relation with symptoms, bronchodilator use and lung function in patients with mild asthma. *Qual Life Res* 1997;6:353–61.
31. Peto V, Jenkinson C, Fitzpatrick R, *et al*. Measuring mental health in amyotrophic lateral sclerosis (ALS): a comparison of the sf-36 mental health index with the psychological general well-being index. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2001;2:197–201.
32. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Qual Life Res* 1993;2:367–8.
33. Stewart AL, Ware JE. *Measuring functioning and well-being: the medical outcomes study approach*. Durham: Duke Univ, 1992.
34. Sloan JA, Aaronson N, Cappelleri JC, *et al*. Assessing the clinical significance of single items relative to summated scores. *Mayo Clin Proc* 2002;77:479–87.
35. de Boer AG, van Lanschot JJ, Stalmeier PF, *et al*. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? *Qual Life Res* 2004;13:311–20.
36. Terwee CB, Prinsen CAC, Chiarotto A, *et al*. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 2018;27:1159–70.