BMJ Open

Comparison Between Different Scales for Measuring Health-Related Quality of life in a Population Sample of Women and Men, the WHO MONICA Project, Gothenburg, Sweden

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024454
Article Type:	Research
Date Submitted by the Author:	26-May-2018
Complete List of Authors:	Krantz, Emily; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Wide, Ulla; Institute of Medicine, Department of Behavioural and Community Dentistry, Institute for Odontology, The Sahlgrenska Academy, University of Gothenburg Trimpou, Penelope; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Bryman, Inger; Institute for Clinical Sciences, Obstetrics and Gynaecology at The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden Landin-Wilhelmsen, Kerstin; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy
Keywords:	Health Related Quality of Life, Methodology, Population study, Empirical comparison, Questionnaires

Comparison Between Different Scales for Measuring Health-Related Quality of life in a Population Sample of Women and Men, the WHO MONICA Project, Gothenburg, Sweden

Emily Krantz M.D[1], Ulla Wide Reg. Psychologist, Ph.D[2], Penelope Trimpou M.D, Ph.D[3], Inger Bryman M.D, Ph.D[4], Kerstin Landin-Wilhelmsen M.D, Ph.D, Prof[3]

1. Clinic of Internal Medicine, Södra Älvsborgs Hospital, Borås, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ORCID ID: 0000-0002-2639-8980
2. Department of Behavioural and Community Dentistry, Institute of Odontology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
3. Section for Endocrinology, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
4. Department of Reproductive Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute for Clinical Sciences, Obstetrics and Gynaecology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

Emily Krantz (née Amundson) (corresponding author)
e-mail: emily.amundson@vgregion.se
Address: Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy, University of Gothenburg, P.O. Box 428, SE-405 30 Gothenburg, Sweden

Word count: 3918

ABSTRACT

Objective There is a need for empirical comparative studies of Health-related Quality of Life assessment instruments in clinical and population samples. We aim to describe the psychometric properties and compare the results of three generic HRQoL instruments and self-rated health in a population sample.

Design An observational cross-sectional study.

Setting A randomly selected population-based sample from Gothenburg, Sweden, was studied in 2007-2008 in the World Health Organization MONItoring of trends and determinants for CArdiovascular disease (WHO MONICA-GOT).

Participants 414 subjects were included, 77% women, age range 39-77 years.

Interventions HRQoL was measured using 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.

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 between the sub-scales that were qualitatively similar were significant ($p < 0.01$). All sub-scales could differentiate the presence of self-rated ill-health according to Self-Rated Health scale ($p < 0.001$). Self-Rated Health scale correlated strongly with all of the three Health-related Quality of Life scales used.

Conclusions There was a high concordance between the instruments within each domain that was qualitatively similar. All three Health-related Quality of Life instruments (PGWB, SF-36 and NHP) could discriminate the presence of self-rated ill-health. Self-Rated Health scale correlated strongly with PGWB, SF-36 and NHP. This supports the strong association between self-rated health and Health-related Quality of Life in the general population.

ARTICLE SUMMARY

Strengths and limitations of this study

- This study reports the results from the general population of more than 400 Swedish subjects with a largely complete data set where very few comparative studies of Health-related Quality of Life measurements have been published on population samples.
- 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 Self-Rated Health scale 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 a Health-related Quality of Life instrument.
- The study population is comprised of middle-aged and elderly subjects and is not evenly distributed between the sexes which affects the generalizability of the sample.

KEYWORDS
Health Related Quality of Life; Methodology; Population study; Empirical comparison; Questionnaires.

INTRODUCTION

Health-Related Quality of life (HRQoL) is an important variable in clinical practice and in medical literature with significant consequences for patients and for society. As the general population ages and as 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. There is a growing number of HRQoL measurement instruments available to researchers and their sophistication, variety and scope is increasing - making it difficult to know which one to choose when planning a study. The definition of HRQoL, and how it is best measured, is the subject of continuing debate and indeed, critique [1-3]. Yet, few studies apply different instruments and compare the results, and even fewer do so in general population samples.

Studies done in Dutch population samples in 1996 [4] and 1997 [5] and in a Brazilian population sample in 2011[6] are some examples. All three 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 conclude that it is important to define one's research question and underline 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.

In this study, the aim is to meet the need for empirical comparative studies of HRQoL assessment instruments by describing the psychometric properties and comparing the results of three different, widely used, generic (not disease specific) HRQoL instruments in a population sample of men and women. A further aim was to evaluate the sub-scales of the different instruments that measure the same domain. Finally, the association between the HRQoL measures and a single-item measure of Self-Rated Health was assessed.

METHODS

Study setting

This is an observational cross-sectional study of a randomly selected population-based sample, n=414, from Gothenburg, Sweden, the World Health Organization (WHO), MONItoring of trends and determinants for CArdiovascular disease (MONICA-GOT) study 2007-2008.

Sample selection process

In 1995, 2400 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 [7]. In all, 1616 individuals were included. The non-attenders in 1995 could not participate due to travel, living abroad, unwillingness to attend, or inability to attend due to illness of a relative. The participants were examined at a medical clinic. Hormone and bone measurements were taken at random on one in four of all of the male participants and all of the female participants in the age group 25–44, and on *all* females aged 45–64. In total, measurements were taken on 662 participants (70% women). This subset was invited for re-evaluation in 2007-08 [8]. Of these 662 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 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 39-77 years).

Procedure

The participants completed the questionnaires while visiting the Center for Endocrinology and Metabolism at the Sahlgrenska University hospital, Gothenburg for medical examinations in the re-examination of the WHO MONICA study. After blood sampling in the fasting state between 8-9 am, all participants received breakfast with coffee or tea during which the questionnaires were administered to all subjects in the following order: Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB), Medical Outcomes Study Short-Form 36 (SF-36), and a single item Self-Rated Health scale (SRH). A single operator performed the measurements and administrations on all subjects. No personal guidance was given except for the instructions.

Measurements

The Nottingham Health Profile (NHP)

NHP measures aspects of subjective health using a two-part questionnaire [9]. 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 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, i.e. 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 [10].

The Psychological General Well-being Index (PGWB)

The PGWB was designed to measure personal affective or emotional states reflecting a sense of well-being or distress intended for use in community surveys [11 12]. The PGWB includes 22 items, with a six-grade Likert style response format where a high score represents a better HRQoL. The scores are summarized into an overall well-being score (PGWB Total score, range 22-132), and also divided into six sub-scales: 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). PGWB is not as commonly used as the other scales but it has been used in clinical trials and has performed well in both population-based and mental health samples [13].

The Medical Outcomes Study Short-Form 36 questionnaire (SF-36)

The SF-36 is a multipurpose health survey comprised of 36 items where a high score represents a better HRQoL [14]. It yields an eight-scale profile of functional health and well-being scores: 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 Physical Component Summary. The Mental Component Summary is comprised of the sub-scales for Vitality, Social Functioning, Role Emotional, and Mental Health, whereas the Physical Component Summary is comprised of the sub-scales 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 [15].

Self-rated Health (SRH)

Self-rated health was measured with a single question: participants were asked to rate their current health status between 0 and 100 on a linear analogue self-assessment scale; 0 being the worst conceivable and level and 100 the best conceivable level. The item is identical to question number 6 published in the 1990 edition of EQ-5D [16]. Single-item health indicators, such as the one used here, have consistently been shown to be strong correlates of objective health and even as predictors of mortality [17-19].

Background variables

Age in whole years and gender 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 of each the instrument's sub-scales including mean, median, standard deviation (SD), percentage of patients with lowest (floor effect) and highest (ceiling effect) possible scores were calculated. The non-parametric Mann-Whitney U test was used to compare bivariate variables, since the results were not normally distributed. Internal consistency was examined using Cronbach's Alpha, $\alpha > 0.70$ was considered significant. Correlation analyses between the instruments were focused on comparing the dimensions that were qualitatively similar, see Table 1. Comparisons between similar sub-scales in the SF-36 and the NHP are similar to ones made in early studies [6 20-23]. 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 $r_s \geq 0.50$. Regression analysis using the R^2 coefficient of determination was also calculated for certain sub-scale comparisons. The presence of self-rated ill-health was defined using the SRH item split at the median and the discriminative ability of all the sub-scales was tested using Mann-Whitney U test.

All statistical analyses were calculated using Statistical Package for the Social Sciences (SPSS v. 24) software. A p-value of <0.01 was chosen to reduce the risk for significance by chance due to multiple tests (type II error). SF-36 scores were calculated using scoring software obtained from Optum™ license number QM03712, Mental and Physical component scores were calculated using 1998 US norms. NHP scores were reversed for consistency with the other scales to facilitate comparisons.

Missing values were imputed in NHP questionnaire if < 80% of the values in a given sub-scale were missing. If >80% of the questions were answered in a sub-scale the median value was calculated and imputed. Imputing was considered unnecessary when analysing PGWB and SRH considering the sample size and the relatively few missing answers.

We used the STROBE cross sectional checklist when writing our report [24].

Patient involvement

No patients 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 its participants.

Table 1. Qualitative comparison of Content of Nottingham Health Profile (NHP), Short-Form 36 (SF-36), Psychological General Well-being index (PGWB), and Self-Rated health.

Domain name	NHP	SF-36	PGWB	Self-Rated Health
Social Functioning	Social Isolation	Social Functioning	-	-
Pain	Pain	Bodily Pain	-	-
Physical Functioning	Physical Mobility	Physical Functioning	-	-
Mental Health	Emotional Reactions	Mental Health	Anxiety & Depressed Mood	-
Vitality	Energy	Vitality	Vitality	
General Health	-	General Health	General Health	Self-Rated Health
Summary Scores	-	Physical Component Summary & Mental Component Summary	PGWB Total score	Self-Rated Health

RESULTS

Characteristics of the subjects

The mean age of the subjects who completed the examination and questionnaires (n = 412) was 62.8 years, range 39-78. 77% were women with a mean age 63.7 years, the men had a mean age 59.6 years ($p < 0.001$). The average number of school years was 12, no statistically significant difference was found between men and women regarding education (data not shown). 3 (1%) of the participants had incomplete or largely incomplete NHP and PGWB questionnaires and 9 (2%) of the participants did not complete SRH. The SF-36 responses were also largely complete with no more than 2.4% missing values in any sub-scale.

Questionnaire scores and distribution

Descriptive statistics for each of the HRQoL instruments are presented for the whole group and for men and women separately in Table 2. Men and women scored similarly in all the NHP sub-scales and in SRH. There was a significant difference between the sexes in the PGWB sub-scales Depressed Mood, General Health and Total score and in the SF-36 sub-scales Physical Functioning Bodily Pain, Mental Health and the Physical Component Summary. Men reported a better HRQoL than the women throughout these scales. No scale had more than 2% missing answers (data not shown).

The distribution of the results was skewed for all of the instruments used, which is to be expected when measuring HRQoL in a general population sample (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 sub-scales. The highest proportion of respondents scoring at the ceiling in the NHP sub-scales was in the sub-scales Social Isolation (84%), Energy (70%) and Physical Mobility (66%). The highest ceiling effects in the SF-36 were seen in the sub-scales Role Emotional (69%), Role Physical (60%), and Social Functioning (59%). The highest proportion of ceiling scores in PGWB was seen in the sub-scale Depressed Mood (42%). SRH was the least skewed of all the scales 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 $\alpha = 0.766$, range 0.656-0.870; PGWB mean $\alpha = 0.854$, range 0.762-0.956; SF-36 mean $\alpha =$

0.857, range 0.831-0.914). Two of the sub-scales in the NHP fell below the standard recommended $\alpha > 0.70$ for group comparisons (Social isolation and Sleep). All of the eight SF-36 sub-scales and four of six sub-scales in the PGWB had α-coefficients >0.80.

Table 2. Descriptive statistics and features of the measures, Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and Self-Rated Health.

	All n=412					Men n = 95		Women n = 317		
	Mean (SD)	Median	Floor (%)	Ceiling (%)	Cronbach's α	Mean (SD)	Median	Mean (SD)	Median	p-value [(a)]
PGWB										
Anxiety	24.1 (5.1)	26	0	34	0.904	24.9 (5.0)	26.5	23.9 (5.1)	25	0.053
Depressed mood	15.9 (2.7)	17	0.2	42	0.880	16.5 (2.5)	18	15.9 (2.7)	17	**0.001**
Positive well-being	16.4 (3.8)	17	0	1.2	0.845	17.1 (3.7)	18	16.3 (3.8)	17	0.039
Self-control	15.5 (2.6)	16	0	18	0.764	15.9 (2.4)	16	15.4 (2.6)	16	0.112
General health	14.3 (3.2)	15	0.2	16	0.762	15.1 (3.0)	16	14.1 (3.1)	15	**0.002**
Vitality	17.3 (4.2)	18	0.5	3	0.867	18.0 (4.2)	19	17.2 (4.1)	18	0.062
Total Score	103.6 (18.6)	109	0	0.7	0.956	107.6 (18.2)	115	102.9 (18.4)	107	**0.003**
NHP[§]										
Emotional Reaction	90.0 (18.8)	100	0.7	63	0.825	90.7 (20.1)	100	89.8 (18.4)	100	0.115
Sleep	78.7 (27.2)	88.9	2.7	44	0.686	84.7 (24.3)	100	76.9 (27.8)	88.9	0.011
Energy	83.3 (30.6)	100	7.2	70	0.769	87.3 (26.2)	100	82.1 (31.7)	100	0.199
Pain	85.8 (25.2)	100	1.9	60	0.870	89.6 (21.7)	100	84.7 (26.1)	100	0.056
Physical Mobility	91.0 (17.0)	100	0.5	66	0.799	91.9 (17.2)	100	90.7 (17.0)	100	0.232
Social Isolation	94.3 (15.3)	100	0.2	84	0.656	95.2 (13.6)	100	94.1 (15.8)	100	0.442
SF-36										
Physical Functioning	78.0 (23.7)	85	0.7	23	0.914	84.8 (23.0)	95	75.9 (23.5)	85	**0.000**
Role Physical	74.5 (37.3)	100	14	60	0.878	82.4 (32.0)	100	72.1 (38.5)	100	0.026
Bodily Pain	69.5 (25.7)	72	1.4	27	0.848	76.9 (24.4)	84	67.3 (25.7)	72	**0.001**
General Health	69.3 (23.4)	72	0.7	7.7	0.831	73.3 (23.0)	77	68.1 (23.4)	72	0.033
Vitality	66.1 (23.6)	70	1	4.6	0.853	70.2 (23.5)	75	64.8 (23.6)	70	0.033
Social Functioning	85.6 (22.7)	100	1.2	59	0.840	88.3 (20.3)	100	84.8 (23.4)	100	0.180
Role Emotional	78.8 (35.3)	100	11	69	0.837	82.8 (33.9)	100	77.6 (35.7)	100	0.123
Mental Health	77.6 (19.9)	84	0.2	9.7	0.853	81.7 (18.9)	88	76.3 (20.0)	80	**0.005**
Physical Component Summary	48.7 (10.3)	51.1	-	-	-	51.4 (9.2)	53.7	47.9 (10.6)	50.4	**0.002**
Mental Component Summary	52.0 (11.2)	56	-	-	-	53.1 (10.8)	57.2	51.6 (11.3)	54.9	0.142
Self-rated Health	75.7 (20.4)	80	0.2	5.3	-	78.3 (18.4)	85	74.8 (20.9)	80	0.205

[(a)]Mann-Whitney-U test. men vs. women $p < 0.01$ considered significant and marked in bold.
§NHP scores are reversed for consistency with the other scales. 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 SRH, SF-36 Mental or Physical Component Summaries and is therefore not shown.

Convergent validity

Correlations between sub-scales in different scales.

Spearman's correlations between the selected sub-scales are shown in Table 3 and correlation coefficients for the similar sub-scales (Table 1) are shown in bold. All inter-scale correlations were significant at $p < 0.01$ in the hypothesized direction.

Correlations between similar domains within NHP and SF-36

The results found in the comparable dimensions of SF-36 and NHP are shown in Figure 2. There were significant positive correlations between all the similar sub-scales of the SF-36 and NHP (Physical Functioning $r_s = 0.72$, Pain $r_s = 0.67$, Vitality $r_s = 0.61$, Social Functioning $r_s = 0.37$, and Mental Health $r_s = 0.70$, all $p < 0.01$). The lowest correlation was in the Social Functioning domain and the highest in the Physical Functioning domain. The correlation-coefficients were strong ($r_s > 0.5$) for all the similar sub-scales except in the Social Functioning domain.

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

The correlations were positive and strong in the sub-scales between PGWB and SF-36 and NHP respectively in the domains they have in common (Mental Health $p < 0.01$, $0.67 \leq r_s \geq 0.74$ and Vitality, $p < 0.01$, $0.59 \leq r_s \geq 0.84$). The PGWB sub-scales were more strongly associated with the SF-36 sub-scales than with the NHP sub-scales. 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 significantly associated with both the SF-36 summary scores but the association was weaker with the Physical Component Summary ($r_s = 0.44$) than with the Mental Component Summary ($r_s = 0.76$).

Correlations between single-item self-rated health and the HRQoL instruments

Correlations between SRH and the General Health sub-scales in the PGWB and the SF-36 were strong ($r_s = 0.66$ and $r_s = 0.77$ respectively). The associations between SRH and the PGWB Total score and the SF-36 Physical Component Summary and Mental Component Summary were also significant and strong. It is notable that there were no weak correlations between SRH and *any* of the other instruments' sub-scales.

Table 3. Spearman's rank correlations between the similar sub-scales (shown in Table 1) in Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and Self-Rated Health.

	PGWB					NHP						SF36							
	Anx	Dep. Mood	General Health	Vital.	Total score	Emot. React	Energy	Pain	Phys. Mobil	Social Isol.	SRH	Phys. Funct	Bodily Pain	Gen. Health	Vital.	Social Funct	Mental Health	PCS[a]	MCS[b]
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	0.434	0.32	0.35	0.42	0.48	0.45	0.27	0.29	1									
Self-rated Health	0.49	0.54	**0.66**	0.63	**0.68**	0.53	0.55	0.46	0.48	0.36	1								
SF-36																			
Phys 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 [a]	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 [b]	0.68	0.69	0.401	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 sub-scales are shown in bold and circled.
† $p > 0.05$, NOT significant. All other correlations are significant $p < 0.01$. §NHP scores are reversed for consistency with the other scales
(a) Physical Component Summary, (b) Mental Component Summary

To further examine and visualize the relationships between some of the sub-scales, we show scatterplot diagrams (Figure 3). In the Social Functioning domain (Figure 2a), the R^2 coefficient is 0.176 NHP vs SF 36 meaning only ca 18% of the variation in social functioning measured with NHP is described by the change in the same dimension measured with SF-36. The General Health domain shows a strong correlation between all three instruments with the highest R^2 coefficient between SRH and SF-36 General health (R^2 = 0.577) (Figure 3c-e). The correlation between PGWB total score and SF-36 Mental Component Summary was also strong and the linear relationship the highest of all the comparisons tested with an R^2 = 0.653 (Figure 3b).

Discriminative ability

To compare the ability of the PGWB, NHP and SF-36 instruments to discriminate subjects on the basis of health, ill health was defined as SRH < 80 (median score). All of the sub-scales could significantly differentiate the presence of self-perceived ill-health ($p < 0.001$) (Table 4).

Table 4. Discriminative ability of Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and Self-Rated Health to identify ill-health.
Mean values for all sub-scales when split according to the Self-Rated Health median score = 80, (0-100 = low-high).

Self-Rated Health split at median	0-79	80-100
PGWB		
Anxiety	21.7	26.0
Depressed Mood	14.5	17.1
Positive Well-Being	14.2	18.1
Self-Control	14.3	16.5
General Health	12.1	16.0
Vitality	14.8	19.2
PGWB Total score	91.3	112.9
NHP§		
Emotional Reaction	80.4	97.1
Sleep	67.2	87.2
Energy	65.7	96.0
Pain	72.1	96.0
Physical Mobility	82.5	97.2
Social Isolation	89.0	98.3
SF-36		
Physical Functioning	63.4	88.8
Role Physical	50.3	92.4
Bodily Pain	53.3	81.7
General Health	51.0	82.9
Vitality	49.6	78.3
Social Functioning	73.5	94.6
Role Emotional	60.4	92.6
Mental Health	65.8	86.2
Physical Component Summary	41.7	53.9
Mental Component Summary	46.0	56.4

$p < 0.001$ for comparisons in all sub-scales, calculated using Mann-Whitney U test
§ NHP scores are reversed for consistency with the other scales

DISCUSSION

The aim of the study was to examine and compare the psychometric properties of three generic Health Related Quality of Life instruments – the NHP, the SF-36 and the PGWB – and their association to Self-Rated Health, when used in a general population sample. The instruments showed acceptable reliability, and the sub-scales measuring the same Health Related Quality of Life domain showed strong associations. The distributions were skewed with considerable ceiling effects. 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 upon the reliability, stability, prominence of ceiling/floor effects and validity [25]. Stability was not tested here because of the cross-sectional nature of this this study, but by the other criteria mentioned, the SF-36 and the PGWB performed equally well and both performed slightly better that 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.

We found strong correlations between the PGWB and 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 done in patients with asthma also found a high correlation between the SF 36 Mental Component Summary to the PGWB Total score [26]. They concluded that administering PGWB together with the SF 36 would be redundant. Another study that compared the PGWB and SF-36, that also focused solely on the Mental Health domain has been done in patients with Amyotrophic Lateral Sclerosis [27]. They found that the internal consistency was equivalent between the two instruments and that all the sub-scales in the PGWB correlated strongly with the Mental Health sub-scale in the SF-36. In the present study, a more nuanced approach was taken when comparing the PGWB to the SF-36 by comparing several similar sub-scales and not singling out Mental Health. The results presented here support earlier recommendations to choose one or the other when the goal is to assess mental health. We suggest that the same could also apply when the goal is to assess General Health or Vitality even when using the instruments 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 a

HRQoL instrument [28]. It does, nonetheless, contain aspects of positive well-being that the others may miss [13]. These results are important considering that this is the only study, to our knowledge, that compares the PGWB to the SF-36 empirically in a general population sample and that there is a lack of published validity studies for the PGWB [10].

Our comparison of the SF-36 to NHP showed that 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 [4 5] and in studies done in patients with pulmonary disease [23], cardiovascular disease [21 29], chronic neuropathic pain [22] and haemodialysis [20]. The congruity between the two instruments was the weakest in the Social isolation domain. The Social Functioning sub-scale of the SF-36 was more strongly associated with the NHP sub-scales for Emotional Reactions and Energy - much like the findings of Prieto et al. and Meyer-Rosberg et al [22 23]. Unlike the aforementioned studies, the Social Isolation sub-scale in the NHP was only associated with the SF-36 Mental Health sub-scale and not with Vitality as the others showed. The items in the Social Functioning domain differ qualitatively in their intent which may explain this result [30]. The SF-36 includes two questions asking the respondent how/if their physical and/or mental problems affect their social interactions. The NHP asks about loneliness, social interactions, close friends, and a feeling of being a burden to others in five questions, without the specific connection to physical or mental symptoms. Notably, for the four remaining common domains, which deal with both mental and physical aspects in HRQoL, each pair of NHP and SF-36 dimensions were strongly correlated. Both the SF-36 and the NHP had the same ability to discriminate the presence of self-rated ill health.

The only previous population based sample, to our knowledge, between the SF-36 and the NHP was done by Faria et al. in 40 community dwelling subjects in Brazil with a mean age of 70 years in 2011 [6]. The results were similar regarding internal consistency and convergent validity, with the exception of the correlation in the Vitality domain between the two instruments. We found a strong association between the two, whereas Faria et al found no significant correlation. 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. The differences between the SF-36 and NHP that were seen in this study were minimal. Like Prieto et al, who studied patients with lung disease, we conclude that it is questionable whether these small differences are clinically relevant even when the instruments are applies in a population sample [23].

In this study, the measure of self-rated health (SRH) correlated significantly, not only with similar sub-scales in the General Health domain, but also with *all* the other instruments' sub-scales. On the whole, the correlations between SRH and the NHP were the least pronounced of all the comparisons made, with moderate (yet statistically significant) correlations for the sub-scales Sleep, Pain, Physical Mobility, and Social isolation. As expected, this population sample did not report serious problems within any of these domains using the NHP. This makes it reasonable to conclude that these aspects were not a major cause of distress for the subjects and did not, in turn, affect SRH. A similar effect was suggested by de Boyer et al. regarding the correlation between SRH and Pain before and after surgery in patients with head-and-neck cancer, using the Medical Outcomes Study Short-Form 20 [31]. The broad association of SRH with all the other instruments sub-scales indicate that SRH could be considered a measure of overall HRQoL.

It has been debated whether single question general health items are reliable and relevant when measuring HRQoL. The overall impression in the literature is that they offer a broad-ranging assessment of health and of quality of life [10 19 32]. The results of this study support that SRH is strongly associated with all domains measured when assessing HRQoL [17]. We suggest, therefore, that a single item SRH question is a useful measure of overall quality of life in epidemiological studies when the need for quick and easy administration is pertinent [10 31 32]. However, SRH cannot been seen as a substitute for multi-item questionnaires when more specific information about mental functioning, sleep and anxiety e.g. are required.

The present study has strengths and weaknesses. Very few comparative studies of HRQoL measurements have been published on population samples. This study reports the results from the general population of more than 400 subjects with a largely complete data set. However, the inclusion of a subset of the original study group from 1995 resulted in a study population of middle-aged and elderly subjects which affects the generalizability of the sample even if the follow-up rates were high. The differences between women and men must be interpreted carefully since the sexes were not evenly distributed in the study group. 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 SRH. Another limitation is the cross-sectional design which makes it impossible to report on the responsiveness of the instruments, which is an important criterion when evaluating an HRQoL instrument. 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 minimizing the risk for systematic error.[17]

CONCLUSION

There was a high concordance between the instruments for evaluating Health Related Quality of Life within each domain that was qualitatively similar. PGWB performed as well as SF-36 and better than NHP regarding internal consistency. All three instruments could discriminate the presence of self-rated ill or good health and Self-Rated Health correlated significantly with all the other instruments' sub-scales. The results support the theory that there is a strong association between self-rated health and Health Related Quality of Life.

Author Contributions

EK contributed to study design, database preparation, data analysis, interpretation of the results, and had the main responsibility of writing the paper. UW contributed to data analysis, interpretation of the results, and revising the paper. PT and IB contributed to the study design, data collection and interpretation of the results. KL-W is principal investigator for the WHO MONICA cohort, initiated this collaborative project, defined the research question, contributed to the study design, data collection, the interpretation of the results, and revising the paper.
All authors read, revised and approved the final version of the paper.

Funding This study was supported by grants from The Swedish Heart and Lung Foundation, The Faculty of Medicine at University of Gothenburg, The Gothenburg Medical Association and the Agreement for Medical Education and Research (ALF) at the Sahlgrenska University Hospital.

Conflict of Interest

None of the authors have any competing interests.

Compliance with Ethical Standards

This study complies with local ethical standards.

Ethical approval

This study was approved by the Regional Ethical Review Board Gothenburg (Reg. No. 088/06) and the Swedish National Data Inspection Board.

Informed consent

Each participant was given written information about the aim of the study and all participants provided written, informed consent.

Data Sharing Statement

No additional data is available.

Acknowledgements

The excellent help from biomedical assistant Stella Nakate for coordinating the study is gratefully acknowledged.

References

1. Gill TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *Jama* 1994;272(8):619-26. [published Online First: 1994/08/24]
2. Post MW. Definitions of quality of life: what has happened and how to move on. *Topics in spinal cord injury rehabilitation* 2014;20(3):167-80. doi: 10.1310/sci2003-167 [published Online First: 2014/12/09]
3. Moons P, Budts W, De Geest S. Critique on the conceptualisation of quality of life: a review and evaluation of different conceptual approaches. *International journal of nursing studies* 2006;43(7):891-901. doi: 10.1016/j.ijnurstu.2006.03.015 [published Online First: 2006/05/16]
4. 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996;5(1):165-74. [published Online First: 1996/02/01]
5. 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. *Medical care* 1997;35(5):522-37. [published Online First: 1997/05/01]
6. 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. *Revista brasileira de fisioterapia (Sao Carlos (Sao Paulo, Brazil))* 2011;15(5):399-405. [published Online First: 2011/10/18]
7. Wilhelmsen L, Johansson S, Rosengren A, et al. Risk factors for cardiovascular disease during the period 1985-1995 in Goteborg, Sweden. The GOT-MONICA Project. *Journal of internal medicine* 1997;242(3):199-211. [published Online First: 1997/11/14]
8. Trimpou P, Lindahl A, Lindstedt G, et al. Secular trends in sex hormones and fractures in men and women. *European journal of endocrinology* 2012;166(5):887-95. doi: 10.1530/eje-11-0808 [published Online First: 2012/02/07]
9. Wiklund I. The Nottingham Health Profile--a measure of health-related quality of life. *Scandinavian journal of primary health care Supplement* 1990;1:15-8. [published Online First: 1990/01/01]
10. McDowell I. Measuring Health: a guide to rating scales and questionnaires. 3rd ed. New York: Oxford University Press 2006.
11. Dupuy H. The Psychological General Well-Being (PGWB) Index. In: Wenger N, Mattson M, Furberg C, et al., eds. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq Publishing Inc 1984 170–83.
12. Dupuy H. The Psychological General Well-Being (PGWD) Index. In: N.K. Wenger MEM, C.D. Furburg, J. Elinson (Eds.), ed. Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies. New York: Le Jacq Publishing 1984:170–83.
13. Naughton MJ, Wiklund I. A critical review of dimension-specific measures of health-related quality of life in cross-cultural research. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1993;2(6):397-432. [published Online First: 1993/12/01]
14. Sullivan M, Karlsson J, Ware JE, Jr. The Swedish SF-36 Health Survey--I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Social science & medicine (1982)* 1995;41(10):1349-58. [published Online First: 1995/11/01]

15. 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. *Journal of clinical epidemiology* 1998;51(11):1095-103. [published Online First: 1998/11/17]
16. EuroQol--a new facility for the measurement of health-related quality of life. *Health policy (Amsterdam, Netherlands)* 1990;16(3):199-208. [published Online First: 1990/11/05]
17. Lee S. Self-Rated Health in Health Surveys. In: Johnson TP, ed. Handbook of Health Survey Methods. 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc. 2015.
18. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior* 1997;38(1):21-37. [published Online First: 1997/03/01]
19. Bernhard J, Sullivan M, Hurny C, et al. Clinical relevance of single item quality of life indicators in cancer clinical trials. *British journal of cancer* 2001;84(9):1156-65. doi: 10.1054/bjoc.2001.1785 [published Online First: 2001/05/05]
20. Zengin N, Oren B, Gul A, et al. Assessment of quality of life in haemodialysis patients: a comparison of the Nottingham Health Profile and the Short Form 36. *International journal of nursing practice* 2014;20(2):115-25. doi: 10.1111/ijn.12130 [published Online First: 2014/04/10]
21. Klevsgard R, Froberg 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. *Journal of vascular surgery* 2002;36(2):310-7. [published Online First: 2002/08/10]
22. 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. *European journal of pain (London, England)* 2001;5(4):391-403. doi: 10.1053/eujp.2001.0260 [published Online First: 2001/12/18]
23. 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. *Journal of clinical epidemiology* 1997;50(4):463-73. [published Online First: 1997/04/01]
24. 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. *International journal of surgery (London, England)* 2014;12(12):1495-9. doi: 10.1016/j.ijsu.2014.07.013 [published Online First: 2014/07/22]
25. Polit D. Nursing research: Principles and methods 7th ed. Philadelphia: Lippincott Williams & Wilkins 2004.
26. 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1997;6(4):353-61. [published Online First: 1997/05/01]
27. 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. *Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases* 2001;2(4):197-201. doi: 10.1080/14660820152882205 [published Online First: 2002/04/18]
28. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Quality of life research : an international journal of*

quality of life aspects of treatment, care and rehabilitation 1993;2(6):367-8. [published Online First: 1993/12/01]
29. Falcoz PE, Chocron S, Mercier M, et al. Comparison of the Nottingham Health Profile and the 36-item health survey questionnaires in cardiac surgery. *The Annals of thoracic surgery* 2002;73(4):1222-8. [published Online First: 2002/05/09]
30. Stewart AL, Ware JE. Measuring functioning and well-being : the medical outcomes study approach. Durham: Durham : Duke Univ. Press 1992.
31. 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? *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2004;13(2):311-20. doi: 10.1023/B:QURE.0000018499.64574.1f [published Online First: 2004/04/17]
32. Sloan JA, Aaronson N, Cappelleri JC, et al. Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic proceedings* 2002;77(5):479-87. [published Online First: 2002/05/15]

Legends to Figures

Figure 1

Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psycholgical General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health visual analogue scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

Figure 2

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

Figure 3

Correlation Scatterplots between similar sub-scales in the Nottingham Health Profile (NHP), Psycholgical General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health visual analogue scale (SRH). a) Social functioning SF-36 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs SRH. e) General Health PGWB vs SRH. R^2 = coefficient of deterimination (goodness of fit).

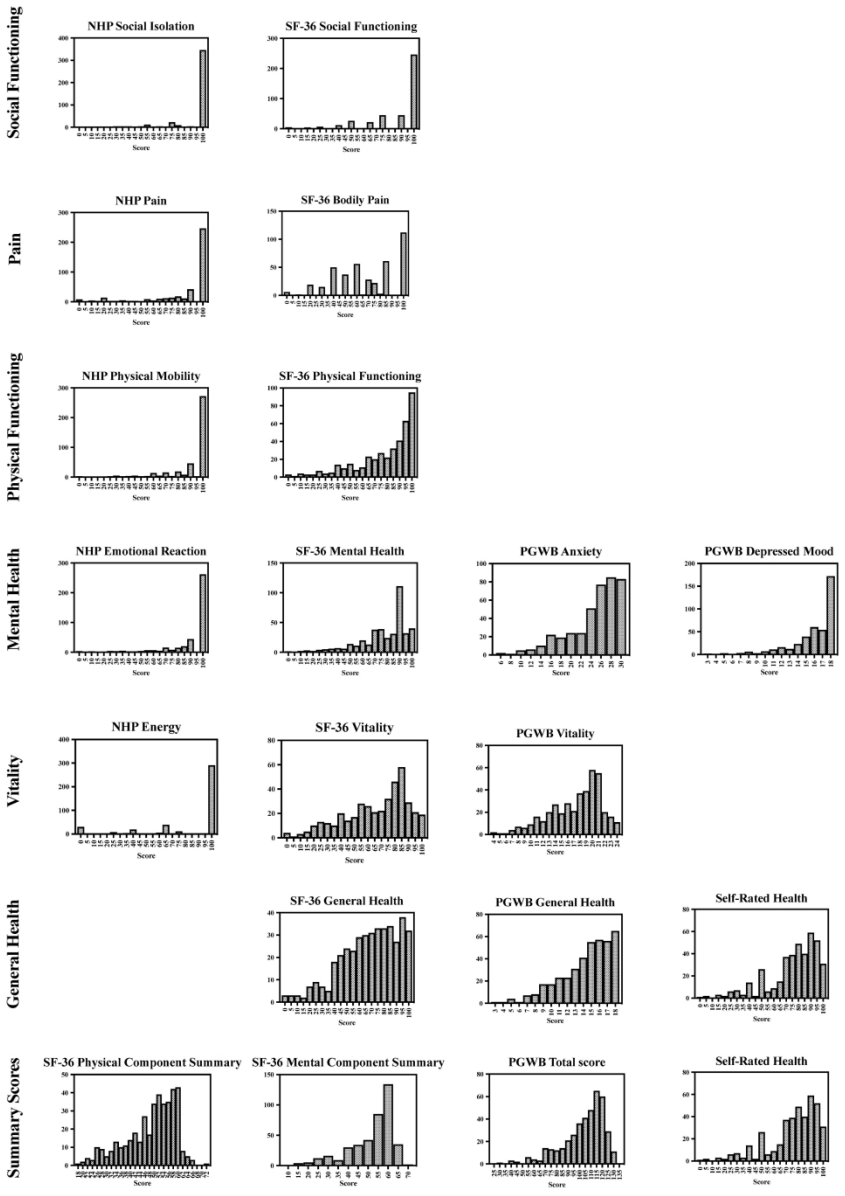


Figure 1 : Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psycholgical General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health visual analogue scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

269x378mm (300 x 300 DPI)

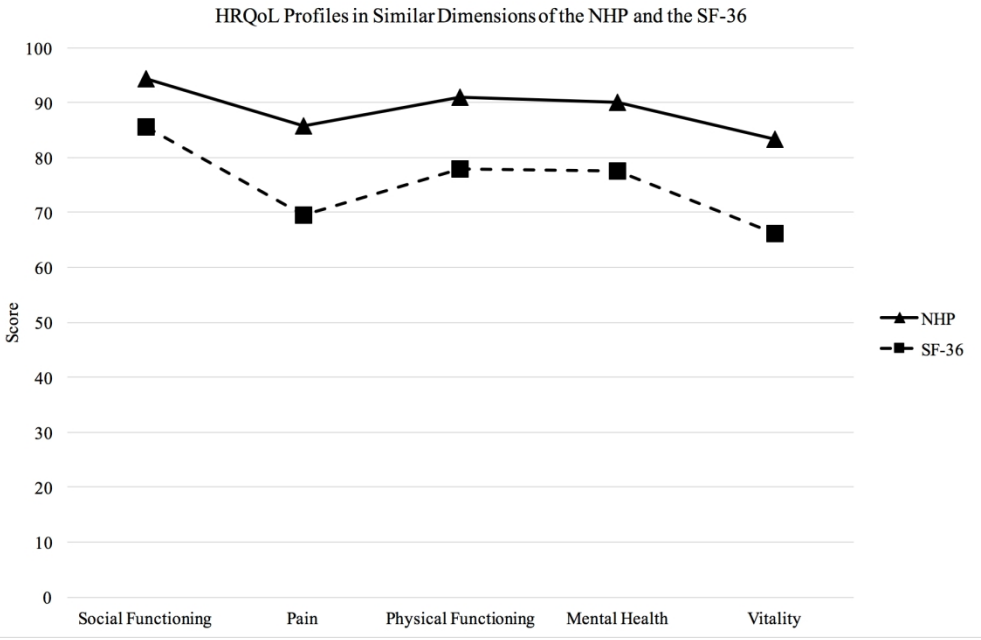


Figure 2. Distributions of scores in comparable dimensions in Nottingham Health Profile (NHP) and Short Form-36 (SF-36). NHP scores are reversed for consistency with the other scales.

344x224mm (144 x 144 DPI)

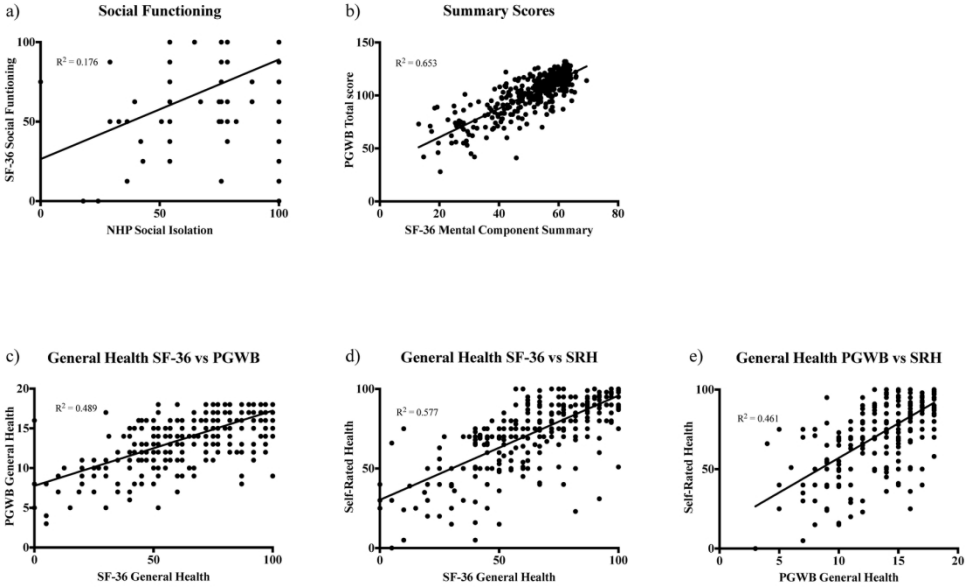


Figure 3: Correlation Scatterplots between similar sub-scales in the Nottingham Health Profile (NHP), Psycholgical General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health visual analogue scale (SRH). a) Social functioning SF-36 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs SRH. e) General Health PGWB vs SRH. R2 = coefficient of deterimination (goodness of fit).

163x98mm (300 x 300 DPI)

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectional reporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	2
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	#3	State specific objectives, including any prespecified hypotheses	4
Study design	#4	Present key elements of study design early in the paper	4
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	5

	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	5-6
Bias	#9	Describe any efforts to address potential sources of bias	7
Study size	#10	Explain how the study size was arrived at	5
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	7
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	7
	#12b	Describe any methods used to examine subgroups and interactions	7
	#12c	Explain how missing data were addressed	7
	#12d	If applicable, describe analytical methods taking account of sampling strategy	n/a
	#12e	Describe any sensitivity analyses	7
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	5
	#13b	Give reasons for non-participation at each stage	5
	#13c	Consider use of a flow diagram	
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	9

	#14b	Indicate number of participants with missing data for each variable of interest	9
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	9
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
	#16b	Report category boundaries when continuous variables were categorized	Table 2
	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Table 3 and 4
Key results	#18	Summarise key results with reference to study objectives	16
Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	19
Generalisability	#21	Discuss the generalisability (external validity) of the study results	18
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

BMJ Open

Comparison Between Different Instruments for Measuring Health-Related Quality of life in a Population Sample of Women and Men, the WHO MONICA Project, Gothenburg, Sweden – an Observational, Cross Sectional Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024454.R1
Article Type:	Research
Date Submitted by the Author:	08-Dec-2018
Complete List of Authors:	Krantz, Emily; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Wide, Ulla; Institute of Medicine, Department of Behavioural and Community Dentistry, Institute for Odontology, The Sahlgrenska Academy, University of Gothenburg Trimpou, Penelope; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Bryman, Inger; Institute for Clinical Sciences, Obstetrics and Gynaecology at The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden Landin-Wilhelmsen, Kerstin; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Mental health
Keywords:	Health Related Quality of Life, Methodology, Population study, Comparison, Instruments

Comparison Between Different Instruments for Measuring Health-Related Quality of life in a Population Sample of Women and Men, the WHO MONICA Project, Gothenburg, Sweden – an Observational, Cross Sectional Study

Emily Krantz M.D[1], Ulla Wide Reg. Psychologist, Ph.D[2], Penelope Trimpou M.D, Ph.D[3], Inger Bryman M.D, Ph.D[4], Kerstin Landin-Wilhelmsen M.D, Ph.D, Prof[3]

1. Clinic of Internal Medicine, Södra Älvsborgs Hospital, Borås, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ORCID ID: 0000-0002-2639-8980
2. Department of Behavioural and Community Dentistry, Institute of Odontology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
3. Section for Endocrinology, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
4. Department of Reproductive Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute for Clinical Sciences, Obstetrics and Gynaecology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

Emily Krantz (née Amundson) (corresponding author)
e-mail: emily.amundson@vgregion.se
Address: Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy, University of Gothenburg, P.O. Box 428, SE-405 30 Gothenburg, Sweden

Word count: 3859

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 are compared. The specific aims were to evaluate the sub-scales 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 2007-2008 in the World Health Organization MONItoring of trends and determinants for CArdiovascular disease (WHO MONICA-GOT).

Participants 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 between the sub-scales that were conceptually similar were significant ($p < 0.01$). All sub-scales could differentiate the presence of self-rated ill-health according to 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.

ARTICLE SUMMARY

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 3 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 a 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.

KEYWORDS
Health-Related Quality of Life; Methodology; Population study; Comparison; Instruments.

INTRODUCTION

Health-Related Quality of life (HRQoL) is an important variable in clinical practice and in medical literature with significant consequences for patients and for society. As the general population ages and as 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 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 upon 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 Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB), 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 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 are compared. A specific aim was to evaluate the sub-scales 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 domains of HRQoL.

METHODS

Study setting

This is an observational cross-sectional study of a population-based sample, n=414, from Gothenburg, Sweden, the World Health Organization (WHO), MONItoring of trends and determinants for CArdiovascular disease (MONICA-GOT) study 2007-2008.

Sample selection process

In 1995, 2400 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 in which 1616 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 illness of a relative. The subjects were examined at a medical clinic. A randomly selected subset of these subjects (every 4th 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]. 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 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

The Nottingham Health Profile (NHP)

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 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, i.e. 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].

The Psychological General Well-Being Index (PGWB)

The PGWB was designed to measure personal affective or emotional states reflecting a sense of well-being or 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 summarized into an overall well-being score (PGWB Total score, range 22-132), and is also divided into six sub-scales: 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].

The Medical Outcomes Study Short-Form 36 questionnaire (SF-36)

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 sub-scales for Vitality, Social Functioning, Role Emotional, and Mental Health, whereas the Physical Component Summary is comprised of the sub-scales 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 and level and 100 the best conceivable level. The item is identical to question number 6 published in the 1990 edition of 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 sub-scales including mean, median, standard deviation (SD), 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 conduct all bivariate analyses, since the results were not normally distributed. Cohen's *d* test was used to calculate the standardized mean effect size between groups, $d > 0.25$ was considered educationally significant, $d > 0.5$ was considered clinically significant [23]. Internal consistency was examined using Cronbach's Alpha, $\alpha > 0.70$ 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 $r_s \geq 0.50$. Regression analysis using the R^2 coefficient of determination was also calculated for certain sub-scale 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 (SPSS 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, 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 sub-scale. 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 6 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 STROBE 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.

Table 1. Comparison of content of 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 & Depressed Mood	-
Vitality	Energy	Vitality	Vitality	
General Health	-	General Health	General Health	Self-rated health
Summary Scores	-	Physical Component Summary & Mental Component Summary	PGWB Total score	Self-rated health

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 percent were women with a mean age 63.7 years, the men had a mean age 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, 5 subjects (1.2%) had incomplete or largely incomplete SF-36 questionnaires, and 9 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 sub-scales and the Self-Rated Health scale. There were statistically significant differences between the sexes in some of the PGWB and SF-36 sub-scales, 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 sub-scales. The highest proportion of respondents scoring at the ceiling in the NHP sub-scales was in the sub-scales Social Isolation (84%), Energy (70%) and Physical Mobility (66%). The highest ceiling effects in the SF-36 were seen in the sub-scales Role Emotional (69%), Role Physical (60%), and Social Functioning (59%). The highest proportion of ceiling scores in PGWB was seen in the sub-scale 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 $\alpha = 0.77$, range 0.66-0.87; PGWB mean $\alpha = 0.85$, range 0.76-0.90; SF-36 mean $\alpha = 0.86$, range 0.83-0.91). Two of the sub-scales in the NHP fell below the standard recommended $\alpha > 0.70$ for group comparisons (Social isolation and Sleep). All of the eight SF-36 sub-scales and four of six sub-scales in the PGWB had α-coefficients >0.80.

Table 2. Descriptive statistics and features of the Psychological 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	-

§NHP scores are reversed for consistency with the other instruments.
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.

Convergent validity

Correlations between the relevant sub-scales (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 sub-scales (Table 1) are shown in bold. All inter-instrument correlations were significant at $p < 0.01$ in the hypothesized 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 sub-scales of the SF-36 and NHP (Physical Functioning $r_s = 0.72$, Pain $r_s = 0.67$, Vitality $r_s = 0.61$, Social Functioning $r_s = 0.37$, and Mental Health $r_s = 0.70$, all $p < 0.01$) (Table 3).

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

The correlations were positive and strong in the sub-scales between the PGWB and the SF-36 and the NHP respectively in the domains they had in common (Mental Health $p < 0.01$, $0.67 \leq r_s \geq 0.74$ and Vitality, $p < 0.01$, $0.59 \leq r_s \geq 0.84$). The PGWB sub-scales were more strongly associated with the SF-36 sub-scales than with the NHP sub-scales. 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 ($r_s = 0.44$) than with the Mental Component Summary ($r_s = 0.76$).

Correlations between the Self-Rated Health scale and the HRQoL instruments

Correlations between the Self-Rated Health scale and the General Health sub-scales in the PGWB and the SF-36 were strong ($r_s = 0.66$ and $r_s = 0.77$ respectively). 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' sub-scales.

Table 3. Spearman's rank correlations between the similar sub-scales (shown in Table 1) in Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and the Self-Rated Health scale (SRHS.)

	PGWB					NHP					SRHS	SF-36							
	Anx	Dep. Mood	General Health	Vitality	Total score	Emot. React	Energy	Pain	Phys. Mobil	Social Isol.		Phys. Funct	Bodily Pain	Gen. Health	Vitality	Social Funct.	Mental Health	PCS[a]	MCS[b]
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	0.434	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																			
Phys. 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 [a]	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 [b]	0.68	0.69	0.401	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 sub-scales are shown in bold and are framed

† $p > 0.05$, NOT significant. All other correlations are significant $p < 0.01$. §NHP scores are reversed for consistency with the other instruments.

(a) Physical Component Summary, (b) Mental Component Summary

Scatterplot diagrams were used to examine and visualize the relationships between some of the sub-scales (Figure 3). The Social Functioning domain (Figure 3a) showed a R^2 coefficient of 0.18 for the NHP vs 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 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 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 sub-scales could significantly differentiate the presence of self-perceived ill-health ($p < 0.001$) and the effect sizes for all sub-scales are above the threshold to be considered clinically significant (Cohen's $d > 0.5$). (Table 4).

Table 4. Discriminative ability of Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and the Self-Rated Health scale to identify ill-health. Mean values are given for all sub-scales 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* [(a)]
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 sub-scales, calculated using Mann-Whitney U test
(a) Cohen's *d* test for standardized mean effect size, $d > 0.5$ considered clinically significant.
§ NHP scores are reversed for consistency with the other instruments.

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 sub-scales 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 upon the reliability, stability, prominence of ceiling/floor effects and validity [29]. Stability was not tested here because of the cross-sectional nature of this this study, but by the other criteria mentioned, the SF-36 and the PGWB performed equally well and both performed slightly better that 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 sub-scales in these two instruments found that internal consistency was equivalent and that all the PGWB sub-scales correlated strongly with the SF-36 Mental Health sub-scale [31]. In the present study, a more nuanced approach was taken by comparing several similar sub-scales 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 a 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 firstly because this is the only study, to our knowledge, that compares the PGWB to the SF-36 applied in a general population sample, and secondly 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 sub-scale was more strongly associated with the NHP sub-scales 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 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 sub-scales in the General Health domain, but with *all* the other instruments' sub-scales. The correlations between Self-Rated Health scale and the NHP were the least pronounced of all the comparisons made, with moderate correlations for the sub-scales 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 e.g. 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 which 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 minimizing 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' sub-scales. 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 Contributions

EK contributed to study design, database preparation, data analysis, interpretation of the results, and had the main responsibility of writing the paper. UW contributed to data analysis,

interpretation of the results, and revising the paper. PT and IB contributed to the study design, data collection and interpretation of the results. KL-W is principal investigator for the WHO MONICA cohort, initiated this collaborative project, defined the research question, contributed to the study design, data collection, the interpretation of the results, and revising the paper.
All authors read, revised and approved the final version of the paper.

Funding This study was supported by grants from The Swedish Heart and Lung Foundation, The Faculty of Medicine at University of Gothenburg, The Gothenburg Medical Association and by grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (Dnr ALFGBG-718611).

Conflict of Interest
None of the authors have any competing interests.

Compliance with Ethical Standards
This study complies with local ethical standards.

Ethical approval
This study was approved by the Regional Ethical Review Board Gothenburg (Reg. No. 088/06) and the Swedish National Data Inspection Board.

Informed consent
Each subject was given written information about the aim of the study and all participants provided written, informed consent.

Data Sharing Statement
Individual participant data that underlie the results reported in this article, after de-identification (text, tables, figures) will be made available. The data will be available beginning 3 months and ending 36 months after publication. Researchers must provide a methodologically sound proposal, and the data may be used to achieve aims in the approved proposal. Proposals should be directed to Emily Krantz (née Amundson) at emily.amundson@vgregion.se. To gain access, data requesters will need to sign a data access agreement.

Acknowledgements
The excellent help from biomedical assistant Stella Nakate for coordinating the study is gratefully acknowledged.

References

1. Parrish RG. Measuring population health outcomes. *Preventing chronic disease* 2010;7(4):A71. [published Online First: 2010/06/17]
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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8 [published Online First: 2010/02/20]
3. Bottomley A, Pe M, Sloan J, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clinical trials (London, England)* 2018;15(6):624-30. doi: 10.1177/1740774518795637 [published Online First: 2018/08/25]
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 internal medicine* 2018 doi: 10.1001/jamainternmed.2018.4710 [published Online First: 2018/10/05]
5. Dodge R, Daly AP, Huyton J, et al. The challenge of defining wellbeing. *International Journal of Wellbeing* 2012;2(3):222-35. doi: 10.5502/ijw.v2.i3.4
6. Schipper H, Clinch JJ, Olweny CLM. Quality of Life Studies: Definitions and Conceptual Issues. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd ed. Philadelphia, Pennsylvvania, USA: Lippincott-Raven 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(7):e022829. doi: 10.1136/bmjopen-2018-022829 [published Online First: 2018/07/25]
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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996;5(1):165-74. [published Online First: 1996/02/01]
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. *Medical care* 1997;35(5):522-37. [published Online First: 1997/05/01]
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. *Revista brasileira de fisioterapia (Sao Carlos (Sao Paulo, Brazil))* 2011;15(5):399-405. [published Online First: 2011/10/18]
11. Wilhelmsen L, Johansson S, Rosengren A, et al. Risk factors for cardiovascular disease during the period 1985-1995 in Goteborg, Sweden. The GOT-MONICA Project. *Journal of internal medicine* 1997;242(3):199-211. [published Online First: 1997/11/14]
12. Trimpou P, Lindahl A, Lindstedt G, et al. Secular trends in sex hormones and fractures in men and women. *European journal of endocrinology* 2012;166(5):887-95. doi: 10.1530/eje-11-0808 [published Online First: 2012/02/07]
13. Wiklund I. The Nottingham Health Profile--a measure of health-related quality of life. *Scandinavian journal of primary health care Supplement* 1990;1:15-8. [published Online First: 1990/01/01]
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, et al., 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1993;2(6):397-432. [published Online First: 1993/12/01]
17. Sullivan M, Karlsson J, Ware JE, Jr. The Swedish SF-36 Health Survey--I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Social science & medicine (1982)* 1995;41(10):1349-58. [published Online First: 1995/11/01]
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. *Journal of clinical epidemiology* 1998;51(11):1095-103. [published Online First: 1998/11/17]
19. EuroQol--a new facility for the measurement of health-related quality of life. *Health policy (Amsterdam, Netherlands)* 1990;16(3):199-208. [published Online First: 1990/11/05]
20. Lee S. Self-Rated Health in Health Surveys. In: Johnson TP, ed. Handbook of Health Survey Methods. 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc. 2015.
21. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior* 1997;38(1):21-37. [published Online First: 1997/03/01]
22. Bernhard J, Sullivan M, Hurny C, et al. Clinical relevance of single item quality of life indicators in cancer clinical trials. *British journal of cancer* 2001;84(9):1156-65. doi: 10.1054/bjoc.2001.1785 [published Online First: 2001/05/05]
23. Wolf FM. Meta-analysis : quantitative methods for research synthesis. Beverly Hills: Beverly Hills : Sage 1986.
24. Zengin N, Oren B, Gul A, et al. Assessment of quality of life in haemodialysis patients: a comparison of the Nottingham Health Profile and the Short Form 36. *International journal of nursing practice* 2014;20(2):115-25. doi: 10.1111/ijn.12130 [published Online First: 2014/04/10]
25. Klevsgard R, Froberg 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. *Journal of vascular surgery* 2002;36(2):310-7. [published Online First: 2002/08/10]
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. *European journal of pain (London, England)* 2001;5(4):391-403. doi: 10.1053/eujp.2001.0260 [published Online First: 2001/12/18]
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. *Journal of clinical epidemiology* 1997;50(4):463-73. [published Online First: 1997/04/01]
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. *International journal of surgery (London, England)* 2014;12(12):1495-9. doi: 10.1016/j.ijsu.2014.07.013 [published Online First: 2014/07/22]
29. Polit D. Nursing research: Principles and methods 7th ed. Philadelphia: Lippincott Williams & 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1997;6(4):353-61. [published Online First: 1997/05/01]
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. *Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases* 2001;2(4):197-201. doi: 10.1080/14660820152882205 [published Online First: 2002/04/18]
32. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1993;2(6):367-8. [published Online First: 1993/12/01]
33. Stewart AL, Ware JE. Measuring functioning and well-being : the medical outcomes study approach. Durham: Durham : Duke Univ. Press 1992.
34. Sloan JA, Aaronson N, Cappelleri JC, et al. Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic proceedings* 2002;77(5):479-87. [published Online First: 2002/05/15]
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? *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2004;13(2):311-20. doi: 10.1023/B:QURE.0000018499.64574.1f [published Online First: 2004/04/17]
36. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0 [published Online First: 2018/03/20]

Legends to Figures

Figure 1

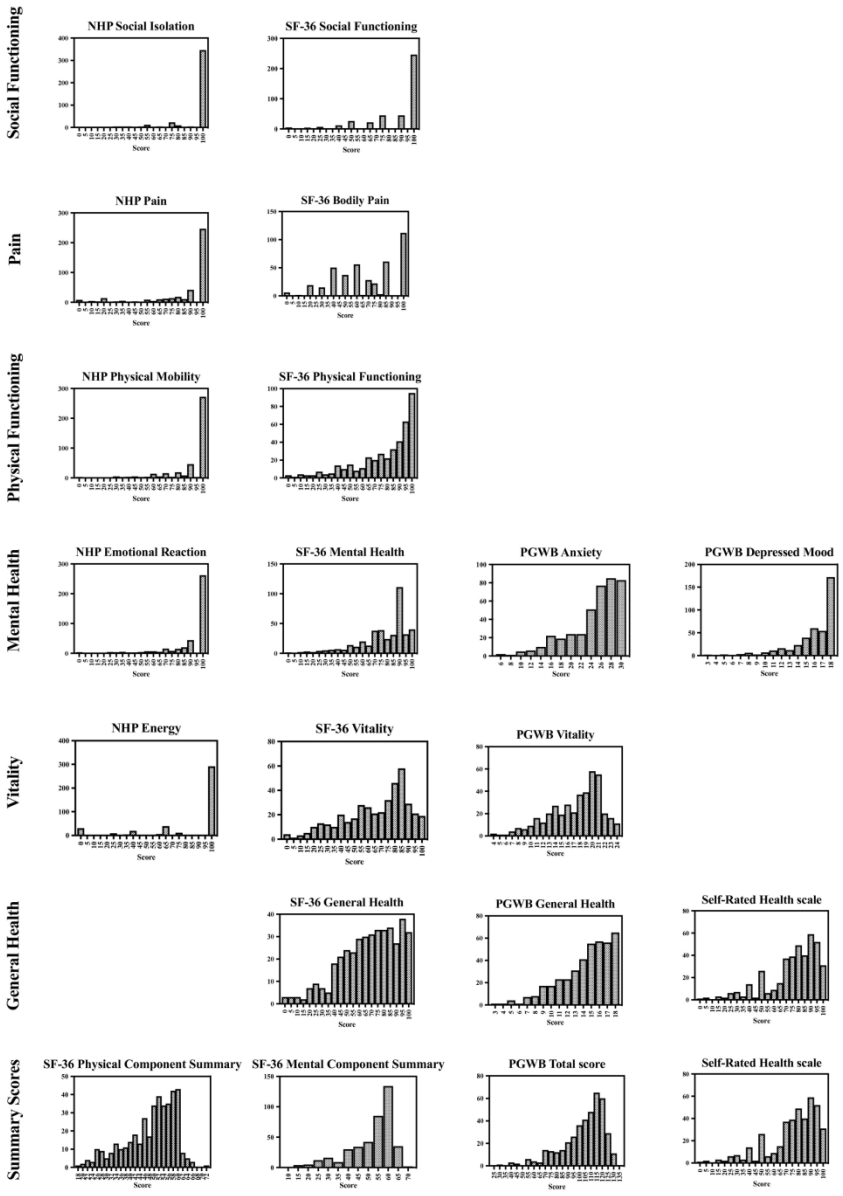
Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

Figure 2

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

Figure 3

Correlation Scatter-plots between similar sub-scales 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 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs Self-Rated Health scale. e) General Health PGWB vs Self-Rated Health Scale. R^2 = coefficient of determination (goodness of fit).



Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

191x269mm (300 x 300 DPI)

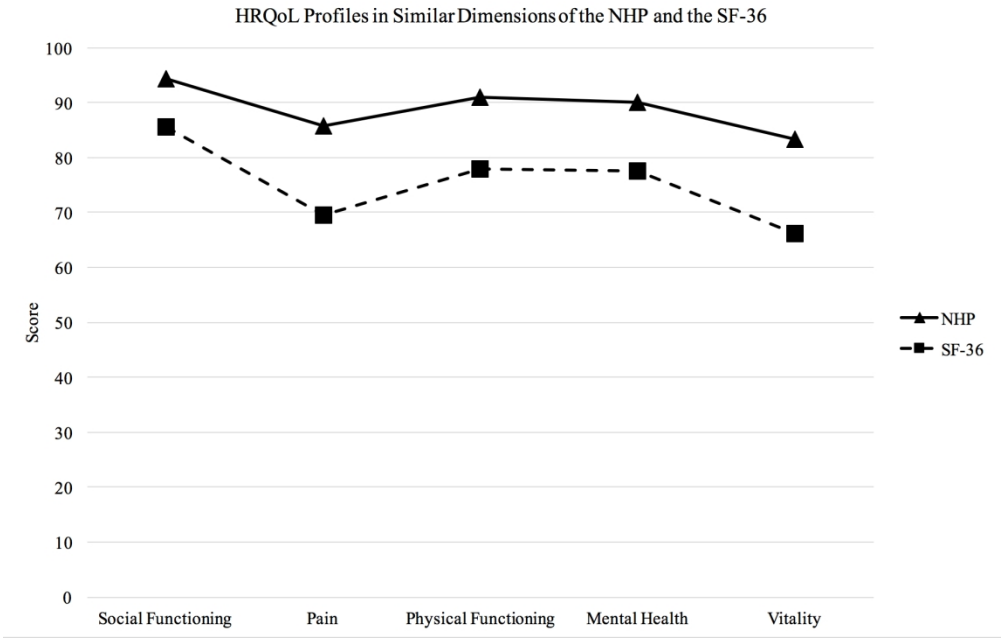


Figure 2
Distributions of scores in comparable dimensions in Nottingham Health Profile (NHP) and Short Form-36 (SF-36). NHP scores are reversed for consistency with the other instruments.

344x224mm (144 x 144 DPI)

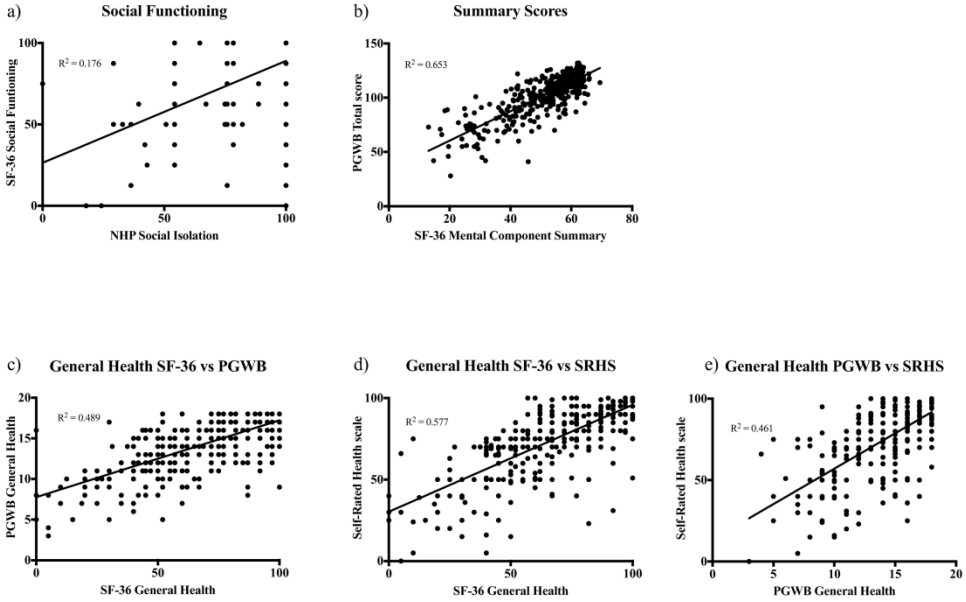


Figure 3
Correlation Scatter-plots between similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and the Self-Rated Health scale (SRHS). a) Social functioning SF-36 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs Self-Rated Health scale. e) General Health PGWB vs Self-Rated Health Scale. R2 = coefficient of determination (goodness of fit).

264x163mm (300 x 300 DPI)

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectional reporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	2
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	#3	State specific objectives, including any prespecified hypotheses	4
Study design	#4	Present key elements of study design early in the paper	4
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	5

	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	5-6
Bias	#9	Describe any efforts to address potential sources of bias	7
Study size	#10	Explain how the study size was arrived at	5
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	7
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	7
	#12b	Describe any methods used to examine subgroups and interactions	7
	#12c	Explain how missing data were addressed	7
	#12d	If applicable, describe analytical methods taking account of sampling strategy	n/a
	#12e	Describe any sensitivity analyses	7
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	5
	#13b	Give reasons for non-participation at each stage	5
	#13c	Consider use of a flow diagram	
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	9

	#14b	Indicate number of participants with missing data for each variable of interest	9
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	9
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
	#16b	Report category boundaries when continuous variables were categorized	Table 2
	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Table 3 and 4
Key results	#18	Summarise key results with reference to study objectives	16
Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	19
Generalisability	#21	Discuss the generalisability (external validity) of the study results	18
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-024454.R2
Article Type:	Research
Date Submitted by the Author:	30-Jan-2019
Complete List of Authors:	Krantz, Emily; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Wide, Ulla; Institute of Medicine, Department of Behavioural and Community Dentistry, Institute for Odontology, The Sahlgrenska Academy, University of Gothenburg Trimpou, Penelope; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy Bryman, Inger; Institute for Clinical Sciences, Obstetrics and Gynaecology at The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden Landin-Wilhelmsen, Kerstin; Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Mental health
Keywords:	Health Related Quality of Life, Methodology, Population study, Comparison, Instruments

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 M.D[1], Ulla Wide Reg. Psychologist, Ph.D [2], Penelope Trimpou M.D, Ph.D [3], Inger Bryman M.D, Ph.D [4], Kerstin Landin-Wilhelmsen M.D, Ph.D, Prof [3]

1. Clinic of Internal Medicine, Södra Älvsborgs Hospital, Borås, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. ORCID ID: 0000-0002-2639-8980
2. Department of Behavioural and Community Dentistry, Institute of Odontology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
3. Section for Endocrinology, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.
4. Department of Reproductive Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden and Institute for Clinical Sciences, Obstetrics and Gynaecology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

Emily Krantz (née Amundson) (corresponding author)
e-mail: emily.amundson@vgregion.se
Address: Institute of Medicine, Sahlgrenska University Hospital at Sahlgrenska Academy, University of Gothenburg, P.O. Box 428, SE-405 30 Gothenburg, Sweden

Word count: 3822

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 sub-scales 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 2007-2008 in the World Health Organization MONItoring of trends and determinants for CArdiovascular disease (WHO MONICA-GOT).

Participants 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 between the sub-scales that were conceptually similar were significant ($p < 0.01$). All sub-scales could differentiate the presence of self-rated ill-health according to 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.

ARTICLE SUMMARY

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 3 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 a 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.

KEYWORDS
Health-Related Quality of Life; Methodology; Population study; Comparison; Instruments.

INTRODUCTION

Health-Related Quality of life (HRQoL) is an important variable in clinical practice and in medical literature with significant consequences for patients and for society. As the general population ages and as 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 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 upon 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 Nottingham Health Profile (NHP), Psychological General Well-Being Index (PGWB), 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 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. A specific aim was to evaluate the sub-scales 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 domains of HRQoL.

METHODS

Study setting

This is an observational cross-sectional study of a population-based sample, n=414, from Gothenburg, Sweden, the World Health Organization (WHO), MONItoring of trends and determinants for CArdiovascular disease (MONICA-GOT) study 2007-2008.

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 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 illness of a relative. The subjects were examined at a medical clinic. A randomly selected subset of these subjects (every 4th 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]. 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 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

The Nottingham Health Profile (NHP)

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 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, i.e. 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].

The Psychological General Well-Being Index (PGWB)

The PGWB was designed to measure personal affective or emotional states reflecting a sense of well-being or 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 summarized into an overall well-being score (PGWB Total score, range 22-132), and is also divided into six sub-scales: 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].

The Medical Outcomes Study Short-Form 36 questionnaire (SF-36)

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 sub-scales for Vitality, Social Functioning, Role Emotional, and Mental Health, whereas the Physical Component Summary is comprised of the sub-scales 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 and level and 100 the best conceivable level. The item is identical to question number 6 published in the 1990 edition of 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 sub-scales including mean, median, standard deviation (SD), 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 conduct all bivariate analyses, 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, $d > 0.5$ was considered clinically significant [23]. Internal consistency was examined using Cronbach's Alpha, $\alpha > 0.70$ 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 $r_s \geq 0.50$. Regression analysis using the R^2 coefficient of determination was also calculated for certain sub-scale 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 (SPSS 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, 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 sub-scale. 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 6 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 STROBE 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.

Table 1. Comparison of content of 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 & Depressed Mood	-
Vitality	Energy	Vitality	Vitality	
General Health	-	General Health	General Health	Self-rated health
Summary Scores	-	Physical Component Summary & Mental Component Summary	PGWB Total score	Self-rated health

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 percent were women with a mean age 63.7 years, the men had a mean age 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, 5 subjects (1.2%) had incomplete or largely incomplete SF-36 questionnaires, and 9 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 sub-scales and the Self-Rated Health scale. There were statistically significant differences between the sexes in some of the PGWB and SF-36 sub-scales, 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 sub-scales. The highest proportion of respondents scoring at the ceiling in the NHP sub-scales was in the sub-scales Social Isolation (84%), Energy (70%) and Physical Mobility (66%). The highest ceiling effects in the SF-36 were seen in the sub-scales Role Emotional (69%), Role Physical (60%), and Social Functioning (59%). The highest proportion of ceiling scores in PGWB was seen in the sub-scale 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 $\alpha = 0.77$, range 0.66-0.87; PGWB mean $\alpha = 0.85$, range 0.76-0.90; SF-36 mean $\alpha = 0.86$, range 0.83-0.91). Two of the sub-scales in the NHP fell below the standard recommended $\alpha > 0.70$ for group comparisons (Social isolation and Sleep). All of the eight SF-36 sub-scales and four of six sub-scales in the PGWB had α-coefficients >0.80.

Table 2. Descriptive statistics and features of the Psychological 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	-

§NHP scores are reversed for consistency with the other instruments.
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.

Convergent validity

Correlations between the relevant sub-scales (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 sub-scales (Table 1) are shown in bold. All inter-instrument correlations were significant at $p < 0.01$ in the hypothesized 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 sub-scales of the SF-36 and NHP (Physical Functioning, Pain, Vitality, Social Functioning, and Mental Health, all $p < 0.01$) (Table 3).

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

The correlations were positive and strong in the sub-scales between the PGWB and the SF-36 and the NHP respectively in the domains they had in common (Mental Health $p < 0.01$, and Vitality, $p < 0.01$). The PGWB sub-scales were more strongly associated with the SF-36 sub-scales than with the NHP sub-scales. 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 sub-scales 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' sub-scales.

Table 3. Spearman's rank correlations between the similar sub-scales (shown in Table 1) in Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and the Self-Rated Health scale (SRHS.)

	PGWB					NHP					SRHS	SF-36							
	Anx	Dep. Mood	General Health	Vitality	Total score	Emot. React	Energy	Pain	Phys. Mobil	Social Isol.		Phys. Funct	Bodily Pain	Gen. Health	Vitality	Social Funct.	Mental Health	PCS[a]	MCS[b]
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	0.434	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																			
Phys. 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 [a]	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 [b]	0.68	0.69	0.401	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 sub-scales are shown in bold and are framed

† $p > 0.05$, NOT significant. All other correlations are significant $p < 0.01$. §NHP scores are reversed for consistency with the other instruments.

(a) Physical Component Summary, (b) Mental Component Summary

Scatterplot diagrams were used to examine and visualize the relationships between some of the sub-scales (Figure 3). The Social Functioning domain (Figure 3a) showed a R^2 coefficient of 0.18 for the NHP vs 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 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 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 sub-scales could significantly differentiate the presence of self-perceived ill-health ($p < 0.001$) and the effect sizes for all sub-scales were above the threshold to be considered clinically significant (Cohen's $d > 0.5$) (Table 4).

Table 4. Discriminative ability of Psychological Well-being Index (PGWB), Nottingham Health Profile (NHP), Short-Form 36 (SF-36), and the Self-Rated Health scale to identify ill-health. Mean values are given for all sub-scales 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* [(a)]
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 sub-scales, calculated using Mann-Whitney U test
(a) Cohen's *d* test for standardized mean effect size, $d > 0.5$ considered clinically significant.
§ NHP scores are reversed for consistency with the other instruments.

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 sub-scales 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 upon the reliability, stability, prominence of ceiling/floor effects and validity [29]. Stability was not tested here because of the cross-sectional nature of this this study, but by the other criteria mentioned, the SF-36 and the PGWB performed equally well and both performed slightly better that 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 sub-scales in these two instruments found that internal consistency was equivalent and that all the PGWB sub-scales correlated strongly with the SF-36 Mental Health sub-scale [31]. In the present study, a more nuanced approach was taken by comparing several similar sub-scales 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 a 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 firstly because this is the only study, to our knowledge, that compares the PGWB to the SF-36 applied in a general population sample, and secondly 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 sub-scale was more strongly associated with the NHP sub-scales 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 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 sub-scales in the General Health domain, but with *all* the other instruments' sub-scales. The correlations between Self-Rated Health scale and the NHP were the least pronounced of all the comparisons made, with moderate correlations for the sub-scales 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 e.g. 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 which 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 minimizing 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' sub-scales. 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 Contributions

EK contributed to study design, database preparation, data analysis, interpretation of the results, and had the main responsibility of writing the paper. UW contributed to data analysis,

interpretation of the results, and revising the paper. PT and IB contributed to the study design, data collection and interpretation of the results. KL-W is principal investigator for the WHO MONICA cohort, initiated this collaborative project, defined the research question, contributed to the study design, data collection, the interpretation of the results, and revising the paper.
All authors read, revised and approved the final version of the paper.

Funding This study was supported by grants from The Swedish Heart and Lung Foundation, The Faculty of Medicine at University of Gothenburg, The Gothenburg Medical Association and by grants from the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (Dnr ALFGBG-718611).

Conflict of Interest
None of the authors have any competing interests.

Compliance with Ethical Standards
This study complies with local ethical standards.

Ethical approval
This study was approved by the Regional Ethical Review Board Gothenburg (Reg. No. 088/06) and the Swedish National Data Inspection Board.

Informed consent
Each subject was given written information about the aim of the study and all participants provided written, informed consent.

Data Sharing Statement
Individual participant data that underlie the results reported in this article, after de-identification (text, tables, figures) will be made available. Researchers must provide a methodologically sound proposal, and the data may be used to achieve aims in the approved proposal. Proposals should be directed to Emily Krantz (née Amundson) at emily.amundson@vgregion.se. To gain access, data requesters will need to sign a data access agreement.

Acknowledgements
The excellent help from biomedical assistant Stella Nakate for coordinating the study, is gratefully acknowledged.

References

1. Parrish RG. Measuring population health outcomes. *Preventing chronic disease* 2010;7(4):A71. [published Online First: 2010/06/17]
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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8 [published Online First: 2010/02/20]
3. Bottomley A, Pe M, Sloan J, et al. Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials. *Clinical trials (London, England)* 2018;15(6):624-30. doi: 10.1177/1740774518795637 [published Online First: 2018/08/25]
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 internal medicine* 2018 doi: 10.1001/jamainternmed.2018.4710 [published Online First: 2018/10/05]
5. Dodge R, Daly AP, Huyton J, et al. The challenge of defining wellbeing. *International Journal of Wellbeing* 2012;2(3):222-35. doi: 10.5502/ijw.v2.i3.4
6. Schipper H, Clinch JJ, Olweny CLM. Quality of Life Studies: Definitions and Conceptual Issues. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd ed. Philadelphia, Pennsylvvania, USA: Lippincott-Raven 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(7):e022829. doi: 10.1136/bmjopen-2018-022829 [published Online First: 2018/07/25]
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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1996;5(1):165-74. [published Online First: 1996/02/01]
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. *Medical care* 1997;35(5):522-37. [published Online First: 1997/05/01]
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. *Revista brasileira de fisioterapia (Sao Carlos (Sao Paulo, Brazil))* 2011;15(5):399-405. [published Online First: 2011/10/18]
11. Wilhelmsen L, Johansson S, Rosengren A, et al. Risk factors for cardiovascular disease during the period 1985-1995 in Goteborg, Sweden. The GOT-MONICA Project. *Journal of internal medicine* 1997;242(3):199-211. [published Online First: 1997/11/14]
12. Trimpou P, Lindahl A, Lindstedt G, et al. Secular trends in sex hormones and fractures in men and women. *European journal of endocrinology* 2012;166(5):887-95. doi: 10.1530/eje-11-0808 [published Online First: 2012/02/07]
13. Wiklund I. The Nottingham Health Profile--a measure of health-related quality of life. *Scandinavian journal of primary health care Supplement* 1990;1:15-8. [published Online First: 1990/01/01]
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, et al., 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1993;2(6):397-432. [published Online First: 1993/12/01]
17. Sullivan M, Karlsson J, Ware JE, Jr. The Swedish SF-36 Health Survey--I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. *Social science & medicine (1982)* 1995;41(10):1349-58. [published Online First: 1995/11/01]
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. *Journal of clinical epidemiology* 1998;51(11):1095-103. [published Online First: 1998/11/17]
19. EuroQol--a new facility for the measurement of health-related quality of life. *Health policy (Amsterdam, Netherlands)* 1990;16(3):199-208. [published Online First: 1990/11/05]
20. Lee S. Self-Rated Health in Health Surveys. In: Johnson TP, ed. Handbook of Health Survey Methods. 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc. 2015.
21. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior* 1997;38(1):21-37. [published Online First: 1997/03/01]
22. Bernhard J, Sullivan M, Hurny C, et al. Clinical relevance of single item quality of life indicators in cancer clinical trials. *British journal of cancer* 2001;84(9):1156-65. doi: 10.1054/bjoc.2001.1785 [published Online First: 2001/05/05]
23. Wolf FM. Meta-analysis : quantitative methods for research synthesis. Beverly Hills: Beverly Hills : Sage 1986.
24. Zengin N, Oren B, Gul A, et al. Assessment of quality of life in haemodialysis patients: a comparison of the Nottingham Health Profile and the Short Form 36. *International journal of nursing practice* 2014;20(2):115-25. doi: 10.1111/ijn.12130 [published Online First: 2014/04/10]
25. Klevsgard R, Froberg 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. *Journal of vascular surgery* 2002;36(2):310-7. [published Online First: 2002/08/10]
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. *European journal of pain (London, England)* 2001;5(4):391-403. doi: 10.1053/eujp.2001.0260 [published Online First: 2001/12/18]
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. *Journal of clinical epidemiology* 1997;50(4):463-73. [published Online First: 1997/04/01]
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. *International journal of surgery (London, England)* 2014;12(12):1495-9. doi: 10.1016/j.ijsu.2014.07.013 [published Online First: 2014/07/22]
29. Polit D. Nursing research: Principles and methods 7th ed. Philadelphia: Lippincott Williams & 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. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1997;6(4):353-61. [published Online First: 1997/05/01]
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. *Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases* 2001;2(4):197-201. doi: 10.1080/14660820152882205 [published Online First: 2002/04/18]
32. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 1993;2(6):367-8. [published Online First: 1993/12/01]
33. Stewart AL, Ware JE. Measuring functioning and well-being : the medical outcomes study approach. Durham: Durham : Duke Univ. Press 1992.
34. Sloan JA, Aaronson N, Cappelleri JC, et al. Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic proceedings* 2002;77(5):479-87. [published Online First: 2002/05/15]
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? *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2004;13(2):311-20. doi: 10.1023/B:QURE.0000018499.64574.1f [published Online First: 2004/04/17]
36. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 2018;27(5):1159-70. doi: 10.1007/s11136-018-1829-0 [published Online First: 2018/03/20]

Legends to Figures

Figure 1

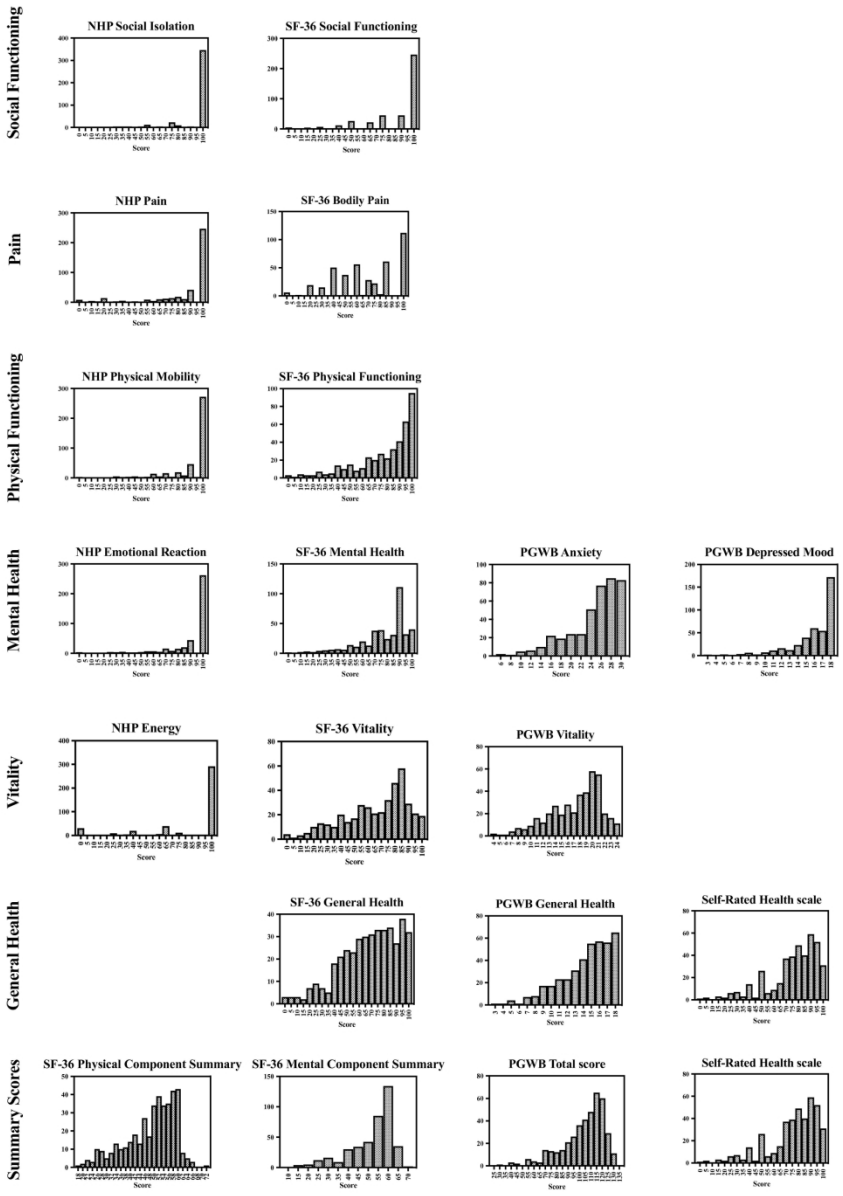
Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

Figure 2

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

Figure 3

Correlation Scatter-plots between similar sub-scales 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 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs Self-Rated Health scale. e) General Health PGWB vs Self-Rated Health Scale. R^2 = coefficient of determination (goodness of fit).



Frequency distributions (histograms) of similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and Self-Rated Health scale. Each row representing a domain (from top to bottom): Social Functioning, Pain, Physical Functioning, Mental Health, Vitality, General Health, Summary Scores.

191x269mm (300 x 300 DPI)

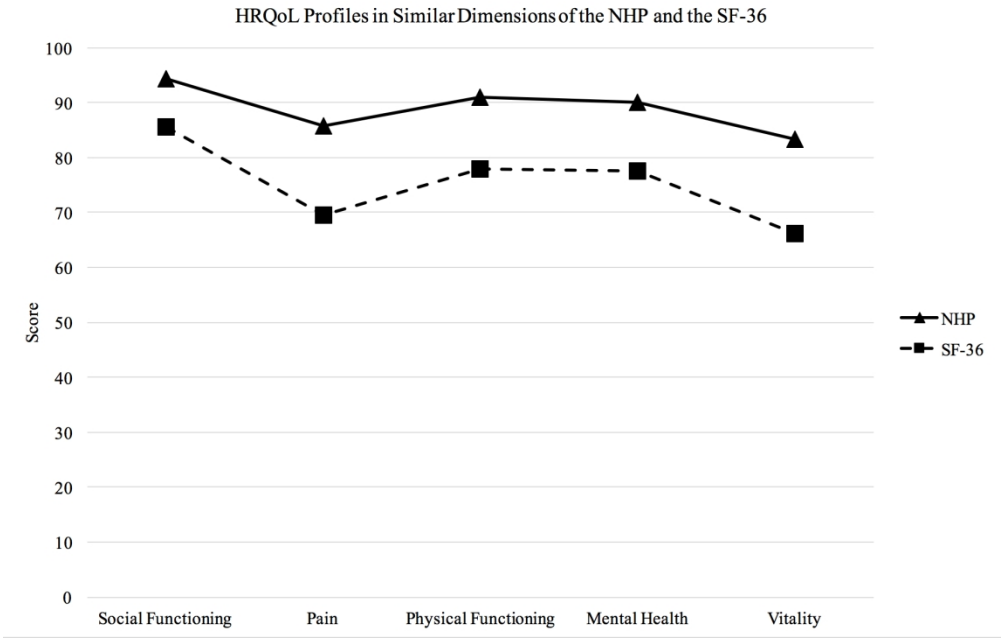


Figure 2
Distributions of scores in comparable dimensions in Nottingham Health Profile (NHP) and Short Form-36 (SF-36). NHP scores are reversed for consistency with the other instruments.

344x224mm (144 x 144 DPI)

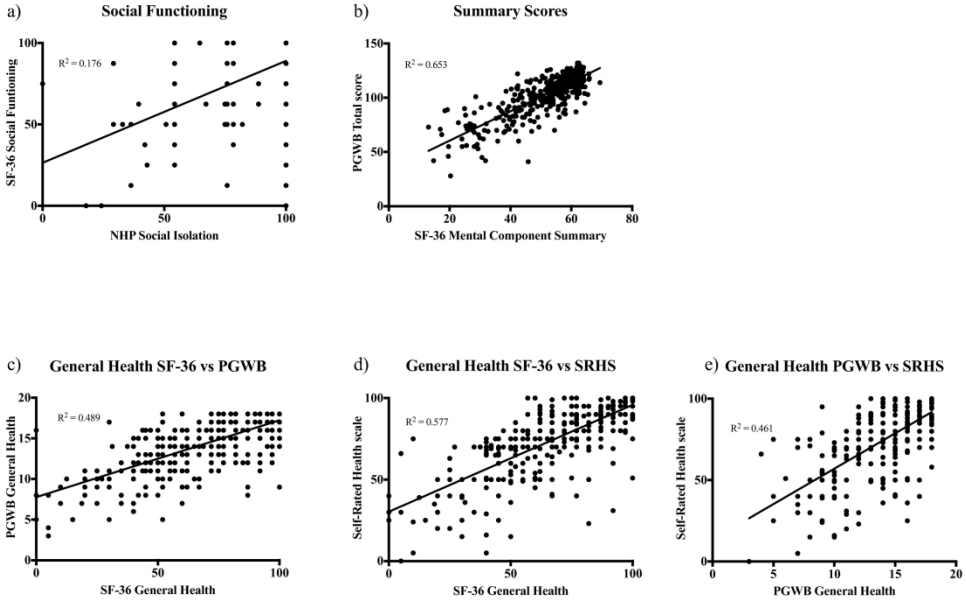


Figure 3
Correlation Scatter-plots between similar sub-scales in the Nottingham Health Profile (NHP), Psychological General Well-Being index (PGWB), Short-Form 36 (SF-36) and the Self-Rated Health scale (SRHS). a) Social functioning SF-36 vs NHP. b) Summary score SF-36 MCS vs PGWB. c) General Health SF-36 vs PGWB. d) General Health SF-36 vs Self-Rated Health scale. e) General Health PGWB vs Self-Rated Health Scale. R2 = coefficient of determination (goodness of fit).

264x163mm (300 x 300 DPI)

Reporting checklist for cross sectional study.

Based on the STROBE cross sectional guidelines.

Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STROBE cross sectional reporting guidelines, and cite them as:

von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies.

		Reporting Item	Page Number
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	1
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	2
Background / rationale	#2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	#3	State specific objectives, including any prespecified hypotheses	4
Study design	#4	Present key elements of study design early in the paper	4
Setting	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4-5
Eligibility criteria	#6a	Give the eligibility criteria, and the sources and methods of selection of participants.	5

	#7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5
Data sources / measurement	#8	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group. Give information separately for for exposed and unexposed groups if applicable.	5-6
Bias	#9	Describe any efforts to address potential sources of bias	7
Study size	#10	Explain how the study size was arrived at	5
Quantitative variables	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	7
Statistical methods	#12a	Describe all statistical methods, including those used to control for confounding	7
	#12b	Describe any methods used to examine subgroups and interactions	7
	#12c	Explain how missing data were addressed	7
	#12d	If applicable, describe analytical methods taking account of sampling strategy	n/a
	#12e	Describe any sensitivity analyses	7
Participants	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable.	5
	#13b	Give reasons for non-participation at each stage	5
	#13c	Consider use of a flow diagram	
Descriptive data	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable.	9

	#14b	Indicate number of participants with missing data for each variable of interest	9
Outcome data	#15	Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable.	9
Main results	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 2
	#16b	Report category boundaries when continuous variables were categorized	Table 2
	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	#17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Table 3 and 4
Key results	#18	Summarise key results with reference to study objectives	16
Limitations	#19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	18
Interpretation	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	19
Generalisability	#21	Discuss the generalisability (external validity) of the study results	18
Funding	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	19