Quality of Life Research (2019) 28:1245–1253
https://doi.org/10.1007/s11136-018-02098-1

Measuring HRQoL following heart valve surgery: the HeartQoL questionnaire is a valid and reliable core heart disease instrument

Charlotte N. Grønset[1] · Lau C. Thygesen[2] · Selina Kikkenborg Berg[3] · Graziella Zangger[4,5] · Marie S. Kristensen[2] · Kirstine L. Sibilitz[6] · Susanne S. Pedersen[7,8] · Neil B. Oldridge[9] · Ann-Dorthe Zwisler[4,5]

Accepted: 29 December 2018 / Published online: 4 January 2019

Abstract

Purpose Patient-reported health-related quality of life is a complementary healthcare outcome and important when assessing treatment efficacy. Using COSMIN methodological recommendations, this study evaluates the validity and reliability of a core heart disease-specific health-related quality of life questionnaire, the HeartQoL questionnaire (Danish version) in a sample of patients following heart valve surgery.

Design This project involved a cross-sectional validity study and a test–retest reliability study.

Methods Eligible patients completed the HeartQoL, the SF-36 health survey questionnaire, and the Hospital Anxiety and Depression Scale following heart valve surgery. Construct validity was tested using a priori hypotheses. Internal consistency reliability was assessed with Cronbach's alpha. An independent sample of patients participated in the test–retest study and reproducibility was determined with relative [intra-class correlation coefficient (ICC)] and absolute reliability [standard error of measurement (SEM) and smallest detectable change (SDC)].

Results Internal consistency was high with Cronbach's alpha ≥ 0.87. ICC was 0.86–0.92. SEM ranged from 0.17 to 0.26 points and SDC ranged from 0.5 to 0.7 points. Construct validity was confirmed with 87% of all a priori hypotheses for predicted variables.

Conclusions The HeartQoL questionnaire demonstrates acceptable construct validity, internal consistency, and test–retest reproducibility in patients following heart valve surgery. Future studies should focus on assessing the responsiveness of the HeartQoL questionnaire over time and following heart valve surgery.

Keywords Heart valve surgery · Health-related quality of life questionnaires · HeartQoL · Measurement properties

✉ Charlotte N. Grønset
charlotte.groenset@regionh.dk

[1] Department of Occupational Therapy and Physiotherapy, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

[2] National Institute of Public Health, University of Southern Denmark, Odense, Denmark

[3] The Heart Centre, Copenhagen University Hospital, Rigshospitalet, Odense, Denmark

[4] The Danish Knowledge Centre for Rehabilitation and Palliative Care (REHPA), Odense University Hospital, Odense, Denmark

[5] University of Southern Denmark, Odense, Denmark

[6] Department of Internal Medicine, Section of Cardiology, Hvidovre University Hospital, Hvidovre, Denmark

[7] Department of Psychology, University of Southern Denmark, Odense, Denmark

[8] Department of Cardiology, Odense University Hospital, Odense, Denmark

[9] College of Health Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Introduction

Heart valve disease is a growing public-health problem due to ageing and the increasing rate of survival of the population [1]. The prevalence of heart valve disease is estimated to be 2.5% [2] with more than 280,000 heart valve surgeries performed worldwide annually [3]. Following heart valve surgery, some patients report anxiety and worries related to readmission and reoperations, postoperative complications, and deconditioning [4] which may prevent or delay return to work and limit activities of daily living [5].

The outcomes after surgery are often evaluated in terms of morbidity and mortality. However, patient-reported outcome measures such as health-related quality of life (HRQoL) are increasingly prominent features in clinical practice, health research, and health policy [6, 7]. Generic HRQoL questionnaires are relevant to broad spectrums of disease or population samples independent of diagnosis. On the other hand, disease-specific instruments are designed to be clinically sensitive to the specific issues of specific patient groups, for example patients with ischaemic heart disease (IHD). The HeartQoL questionnaire [8] was developed in 2014 as an IHD-specific HRQoL instrument. HeartQoL has been validated in patients with angina, myocardial infarction, heart failure, and among patients with stable coronary heart disease, atrial fibrillation, and implantable cardioverter defibrillators [8–12]. Therefore, HeartQoL can be considered a core heart-specific HRQoL questionnaire combining the benefits of both generic and disease-specific questionnaires. To our knowledge, there are no disease-specific HRQoL questionnaires developed for patients with heart valve disease or following heart valve surgery.

The aim of this study is to evaluate the validity and reliability of the HeartQoL questionnaire in a mixed sample of patients following heart valve surgery in accordance with the COSMIN methodological recommendations [13].

Methods

Study design and participants

This project involved two independent studies conducted at different times—a cross-sectional study design aimed at establishing construct validity of HeartQoL and a test–retest reliability study design examined its reliability.

Cross-sectional study

All patients (age ≥ 18) admitted for heart valve surgery in Denmark between January 1st and June 30th 2011 were identified in the Danish National Patient Register (NPR) [14]. The surgeries included conventional open-heart valve replacement surgery alone or combined with coronary artery bypass grafting (CABG) and percutaneous valve replacement or repair. Surgical procedure codes were linked to the National Civil Registration System by the unique personal identification number assigned to all Danish citizens. Information on comorbidity was obtained from the NPR and summarised according to the Charlson comorbidity index (CCI) [15]. Exclusion criteria included address/research protection and emigration. In December 2011, a nationwide survey (the CopenHeart survey [16]) was sent by post to all eligible patients. Patients received the questionnaire package 6–12 months post-surgery which comprised of self-reported HRQoL, mental health, and physical activity questionnaires, questions about cardiac rehabilitation and demographics. A reminder was sent 4 weeks after the initial questionnaire to non-respondents.

Test–retest reliability study

Patients (≥ 18 years) receiving heart valve surgery between November 2013 and May 2014 were eligible to participate. Methodologically, a sample size of a least 50 patients is recommended [17]. A consecutive sample of 150 patients was identified at the Copenhagen University Hospital. The HeartQoL questionnaire (version 2.0; see Patient-reported measures) was sent to patients 6–12 months post-surgery. Following receipt of the first questionnaire, the follow-up HeartQoL was mailed together with supplementary questions regarding time spent to complete the questionnaire and whether major life events had occurred since receiving the initial questionnaire. The term major life event was not specified but kept as an open response category for the respondents to fulfil. Patients not reporting a major life event were considered stable and included in the reliability analysis. A follow-up letter was sent to non-respondents two weeks later.

Patient-reported measures

The HeartQoL, a validated self-administered IHD-specific instrument measuring perceived HRQoL over the previous four weeks [9], includes 14 items, a 10-item physical subscale, and a 4-item emotional subscale with all items forming a global scale. Items are rated on a 4-point Likert scale (range 0–3) with higher scores indicating higher HRQoL. Scores are calculated as the mean of item scores [8, 9]. The HeartQoL questionnaire was translated into Danish according to guidelines for forward–backward translation as a part of the international HeartQoL project [8]. We used two versions of the Danish HeartQoL in this study: HeartQoL version 1.0 in the cross-sectional study and version 2.0 in the test–retest reliability study. Due to one ambiguous item

(item #9) in the original Danish translation (version 1.0), the item was rephrased to improve understanding in version 2.0. The wording of item # 9 "Not feeling relaxed and free of tension?" was changed to "feeling relaxed?" as the Danish double negative confused the meaning.

The Short Form 36 Health Survey (SF-36) is a validated self-administered generic health survey instrument developed to evaluate general health status over the previous four weeks and consists of 36-items, eight dimensions, and two summary scales: a Physical Component Summary (PCS) and a Mental Component Summary (MCS) [18]. An algorithm is used to calculate summary scores that range from 0 to 100, with higher values representing better health status. One item evaluates health transition on which respondents rate their general health (GH) compared to one year earlier [18]. A validated Danish version of SF-36 was used in this study [19].

The Hospital Anxiety and Depression Scale (HADS) is a validated self-rating screening instrument assessing levels of anxiety and depression over the previous week [20, 21]. HADS includes seven items related to symptoms of anxiety (HADS-A) and seven related to symptoms of depression (HADS-D). Scores range from 0 to 21 with higher scores indicating greater symptom severity. A cut-off score ≥ 8 indicates possible or probable clinically relevant anxiety and/or depression [21, 22].

Statistical analyses

For handling missing items, the pairwise deletion method was applied. In all statistical tests, the significance level was set at P value < 0.05. Floor and ceiling effects were considered present when $> 15\%$ of the respondents recorded the lowest (0) and the highest (3) scores, respectively [17]. Data were analysed using SPSS version 22.

Reliability

Analysis of internal consistency reliability was used for each domain and the global scores with Cronbach's alpha values > 0.70 were considered acceptable, > 0.80 good, and > 0.90 excellent [23]. The use of Cronbach alpha as a reliability measure is currently being debated [24, 25] with the COSMIN group claiming Cronbach alpha as the preferred measure for evaluating reliability and the aim is to follow these standards. Paired-sample t tests were used to calculate the significance of the mean difference between HeartQoL test and retest scores. The intra-class correlation coefficient (ICC, $\text{model}_{2,1}$) was used to determine relative reliability [17, 26] with an ICC of ≥ 0.70 recommended as acceptable reliability [23]. Absolute reliability was expressed by standard error of measurement (SEM) calculated as $(SD_{(\text{difference})}/\sqrt{2})$. The smallest detectable change (SDC), which is the magnitude of change necessary to exceed the measurement error and presents the smallest change that can be detected beyond the measurement error, was calculated as $(1.96 \times \sqrt{2} \times \text{SEM})$ [17]. Bland–Altman plots were performed to determine the correlation between the mean scores and the difference between test and retest scores, with a 95% limit of agreement (LoA) ($\text{mean}_{\text{DIFF}} \pm 1.96\text{SD}_{\text{DIFF}}$).

Validity

Construct validity was explored in 15 a priori hypotheses with the PCS, MCS, and GH of the SF-36 and HADS constructs and was considered acceptable with at least 75% of the predefined hypotheses confirmed [27]. Depending on the distribution of the scores, either the Pearson or the Spearman correlation coefficients were used to test for convergent validity with moderate to strong correlations (≥ 0.4) between similar constructs and for discriminant validity with weak correlations (< 0.4) between dissimilar constructs. Known group analysis was used to determine discriminative validity with six *a priori* hypotheses identified for subgroups expected to score differently on the HeartQoL questionnaire (Table 1).

Results

Cross-sectional study

Of the 879 patients admitted for heart valve surgery between January 1st and June 30th 2011, 742 were eligible for this study. The 557 (75%) respondents were significantly more likely to be older (median age, 73 years, range 18–95 years), male (64%), and married (64%) than the 185 non-respondents (Table 2). There was no significant difference between respondents and non-respondents in type of surgery or medical history.

The highest HeartQoL score was reported by 61 respondents (11%) with 224 reporting a score of ≤ 2. Between 2% and 4% of the HeartQoL scores were missing compared to 24% missing SF-36 scores and 10% missing GH and HADS scores (Table 3). All floor and ceiling effects were acceptable, except for the HeartQoL emotional subscale as a ceiling effect was observed in 36% of the respondents (Table 4).

Test–retest reliability study

A total of 150 patients were eligible for the test–retest study. Of these, 92 patients completed both questionnaires (response rate = 61%). Eight respondents were excluded as they had suffered a major life event meaning 84 patients (56%) provided test–retest data. The majority of the respondents were male (64%), median age of 70 years (range 30–88).

Table 1 A priori hypothesis for discriminative validity

A	Gender
	Men will have greater mean scores on HeartQoL, than women
B	SF-36 health transition
	Patients reporting improved health compared with one year ago on the health transition item will demonstrate greater mean scores on the HeartQoL scale, than a patient with no change or deterioration
C	Charlson comorbidity index
	Patients with no or mild comorbidity (CCI ≤ 1) will have greater mean scores on HeartQoL than a patient with severe or very severe comorbidity (CCI ≥ 2)
D	HADS anxiety
	Patients reporting anxiety on HADS-A (scores ≥ 8) will demonstrate lower mean scores on the HeartQoL, than a patient without anxiety
E	HADS depression
	Patients reporting depression on HADS-D (score ≥ 8) will demonstrate lower mean scores on the HeartQoL, than a patient without depression
F	Congestive Heart failure
	Patients with chronic heart failure will have lower mean scores on HeartQoL, than a patient without heart failure

SF-36 short form 36 health survey. *HADS* the Hospital Anxiety and Depression Scale

Table 2 Sociodemographic and clinical characteristics for the study population

Patient characteristics	Total group	Respondents	Non-respondents	*P* value*
Eligible *n* (%)	742	557 (75.1)	185 (24.9)	
Age (years) median, (min/max)	72 (18–95)	73 (18–95)	70 (18–95)	0.003
Gender *n* (%)				0.02
Male	458 (62)	357 (64)	101 (55)	
Marital status *n* (%)				< 0.001
Widow/Widower	164 (22)	120 (22)	44 (24)	
Single	131 (18)	83 (15)	48 (26)	
Married	447 (60)	354 (64)	93 (50)	
Time from surgery to questionnaire *n* (%)				0.64
6–7 months	248 (20)	191 (34)	57 (31)	
8–9 months	242 (33)	181 (32)	61 (33)	
10–12 months	252 (34)	185 (33)	67 (36)	
Type of heart valve surgery *n* (%)				
Mitral valve	166 (22)	126 (23)	40 (22)	0.86
Aortic valve	587 (79)	446 (80)	141 (76)	0.31
Pulmonary valve	11 (1)	6 (1)	5 (3)	0.15
Tricuspid valve	10 (1)	7 (1)	3 (2)	0.77
Percutaneous valve surgery	75 (10)	57 (10)	18 (10)	0.96
Concomitant CABG	206 (28)	154 (28)	52 (28)	0.98
Medical history *n* (%)				
Previous heart valve surgery	4 (1)	3 (1)	1 (1)	1.00
Previous PCI procedure	37 (5)	27 (5)	10 (5)	0.92
Previous CABG procedure	37 (5)	26 (5)	11 (6)	0.77
Comorbid heart failure	162 (22)	113 (20)	49 (26)	0.10
Charlson comorbidity index *n* (%)				0.12
No or mild comorbidity (score ≤ 1)	534 (72)	409 (74)	125 (67)	
Severe or very severe (score ≥ 2)	207 (28)	147 (26)	60 (32)	

CABG coronary artery bypass graft, *PCI* percutaneous coronary intervention, Charlson comorbidity score

*Comparison between non-respondents and respondents: Chi-square or Fischer's exact test

Table 3 Patient-reported outcome scores and the proportion of missing items

Patient-reported instrument	Scores	Missing (%)
HeartQoL [median and (25th–75th quartile)]		
Global, $n=546$	2.29 (1.57–2.79)	2
Physical, $n=546$	2.20 (1.36–2.80)	2
Emotional, $n=534$	2.75 (2.00–3.00)	4
HeartQoL score in tertiles *n* (%)		
Low (≤ 2.00)	224 (41)	
Moderate (2.01–2.99)	261 (48)	
High (=3)	61 (11)	
SF-36 (mean ± SD)		
PCS, $n=422$	44.5 (±10.6)	24
MCS, $n=422$	51.9 (±10.4)	24
GH, $n=504$	65.4 (±22.0)	10
HADS (mean ± SD)		
HADS anxiety, $n=501$	3.5 (±3.6)	10
% anxious Δ	13.6%	
HADS depression, $n=516$	3.3 (±3.5)	7
% depressed Δ	13.8%	

Short form 36 health survey (SF-36; physical (PCS) and mental (MCS) component summary scale and general health (GH) perceptions index score), Hospital Anxiety and Depression Scale (HADS). Δ = HADS score ≥ 8

The mean time interval between test and retest was 13.2 days (SD ± 5.4, range 9–42). The median time to complete the HeartQoL was 5 min (range 5–30.0). Mean differences (± SD) between test and retest ranged from 0.01 (±0.24) to 0.06 (±0.37) ($P>0.05$). $ICC_{2,1}$ values were > 0.80 for each HeartQoL scale with SEM ranging from 0.17 to 0.26 points and SDC_{95} ranging from 0.5 to 0.7 points (Table 4). Bland–Altman plots were conducted and LoA for the HeartQoL global scale was −0.61 to 0.53; the HeartQoL physical scale: −0.79 to 0.67; and the HeartQoL emotional scale: −0.48 to 0.46, see Fig. 1.

Internal consistency reliability

Internal consistency reliability was clearly demonstrated with Cronbach's alpha values ranging from 0.87 to 0.94 for the global scale and each subscale (Table 4).

Validity

Convergent validity was demonstrated with correlations > 0.60 between the HeartQoL global scale and the SF-36 GH, between the physical and emotional subscales and the similar SF-36 PCS and MSC constructs as well as the HADS-A and HADS-D constructs (Table 5). However, for discriminant validity, HeartQoL physical subscale was associated with the dissimilar HADS-D construct ($r=-0.66$) and HeartQoL emotional subscale was associated with the dissimilar SF-36 PCS construct ($r=0.45$) (Table 5). Discriminative validity was confirmed with significant score differences for all *a priori* hypotheses (Table 6).

Table 4 Floor and ceiling effects and reliability parameters of the HeartQoL

	HeartQoL global	HeartQoL physical	HeartQoL emotional
Floor and ceiling effects (%)			
Floor	8.42	13.55	5.61
Ceiling	11.17	13.74	36.00
Internal consistency ($n=557$)			
Cronbach's alpha	0.94	0.93	0.87
Relative and absolute reliability ($n=84$)			
Test (mean ± SD)	2.4 ± 0.6	2.3 ± 0.7	2.6 ± 0.6
Retest (mean ± SD)	2.3 ± 0.6	2.2 ± 0.7	2.6 ± 0.6
Difference test–retest (mean ± SD)	0.04 ± 0.29	0.06 ± 0.37	0.01 ± 0.24
Measurement error			
$ICC_{2,1}$ [95% CI]	0.88 [0.83–0.92]	0.86 [0.79–0.90]	0.92 [0.88–0.95]
SEM	0.21	0.26	0.17
SDC	0.6	0.7	0.5
LoA	− 0.61 to 0.53	− 0.79 to 0.67	− 0.48 to 0.46

$ICC_{2,1}$ intra-class correlation coefficient two-way random model—absolute agreement

CI confidence interval *SEM* standard error of measurement, *SDC* smallest detectable change. *LoA* limit of agreement

Floor effect = minimum score (=0), ceiling effect = maximum score (=3)

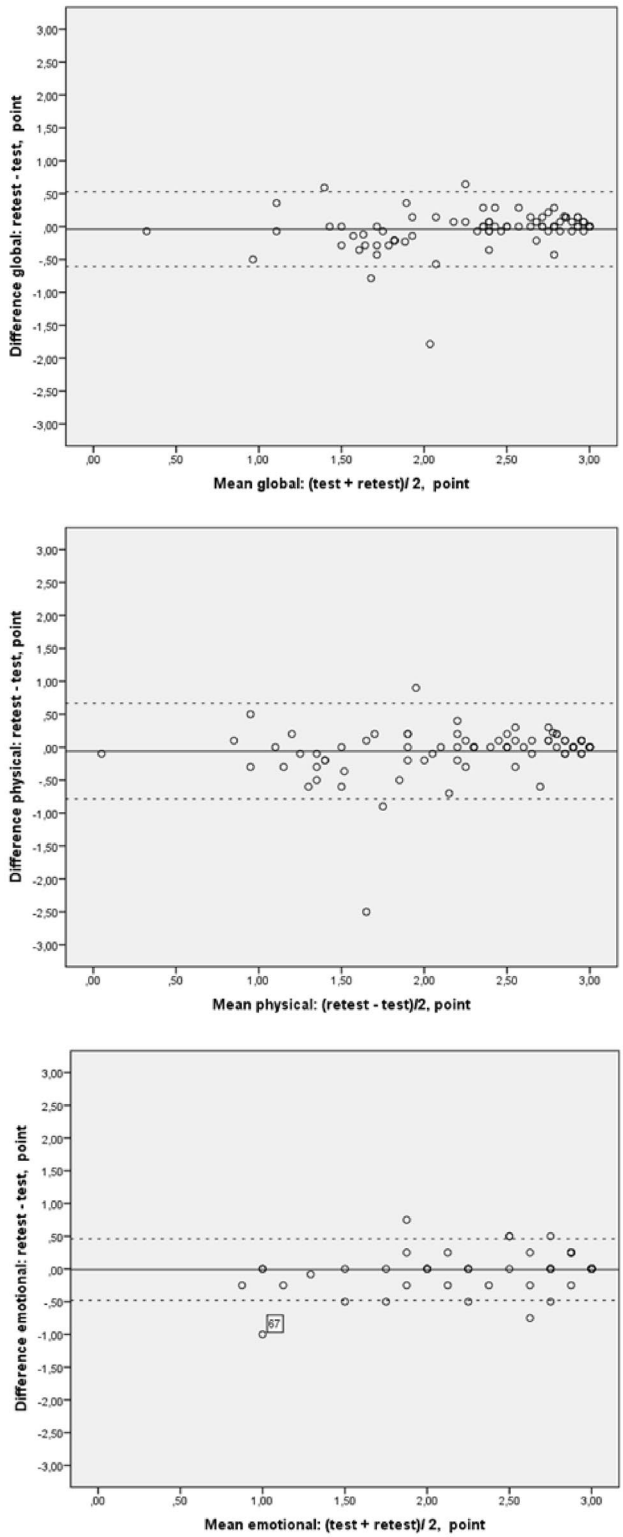

Fig. 1 Bland–Altman plots for the differences between scores from test and retest against the mean of the test and retest for each responder for the HeartQoL global scale, physical, and emotional subscales. The 95% limits of agreement lines are indicated by the dashed lines. The solid lines represent the mean differences scores between the repeated measurements

Discussion

We evaluated the measurement properties of the core IHD-specific HRQoL HeartQoL questionnaire in patients following heart valve surgery and found acceptable construct validity, internal consistency reliability, and test–retest reproducibility although the higher than expected measurement error may be of some concern [17]. However, these results suggest that HeartQoL may be suitable for measuring HRQoL in patients following heart valve surgery. As HeartQoL has been shown to have acceptable reliability and validity in patients with IHD diagnoses of angina pectoris, myocardial infarction, and heart failure [9, 12] as well as in patients with atrial fibrillation [10], the HeartQoL shows promise as a heart disease core HRQoL instrument.

Construct validity analyses supported 13 of the 15 a priori stated hypotheses (87%). Convergent validity of the HeartQoL was confirmed by its strong correlation with SF-36 and HADS across similar constructs. However, discriminant validity correlations were higher than expected between dissimilar constructs, i.e. HeartQoL emotional subscale and SF-36 PCS and between HeartQoL physical subscale and HADS-D. Similar correlations between physical and emotional constructions have been seen in validation studies of MacNew heart disease HRQoL questionnaire [10, 28–30]. As in the MacNew questionnaire, some of the HeartQoL physical items refer to the experience of symptoms and not to physical performance limitations. For example, the formulation of item #8 on the physical scale as "feeling tired, fatigued, low on energy" may lead the respondents to identify with their emotional state rather than with their physical condition. Further, HADS might not be the best reference instrument to hold against the constructs of the HeartQoL, as HADS does not provide an overall score and even more importantly HADS does not measure the same overall construct—HRQoL. HeartQoL clearly discriminates between groups as hypothesised with significant differences between gender, health status, comorbidity, anxiety, depression, and heart failure.

With Cronbach's alpha values > 0.86, internal consistency reliability was confirmed for HeartQoL global scale and both subscales, consistent with the original HeartQoL [9, 10] and the atrial fibrillation [9, 10] validation studies. The lower HeartQoL emotional subscale alpha value may be explained by the small number of items.

A critical aspect for the test–retest is the time interval between tests. In this study, the interval had a relative wide time range, and this may cause both carryover effect for shorter interval and the possibility of response shift for longer interval [31]. Even though there are no standard rules for the interval length between the two tests, a

Table 5 Correlation matrix between HeartQoL questionnaire, the short form 36 health survey physical and mental component summary scales and the General health perception index, and the hospital anxiety and depression questionnaire

		HeartQoL global	HeartQoL physical	HeartQoL emotional
SF36				
A correlation of 0.4 or greater between physical component summary scale (PCS)		Convergent validity	*0.84*	–
A correlation of 0.4 or greater between mental component summary scale (MCS)		–	–	*0.68*
A correlation of 0.4 or greater between General Health Perception index		*0.71*	–	–
A correlation of 0.4 or less between physical component summary scale (PCS)		Discriminant validity	–	0.45
A correlation of 0.4 or less between mental component summary scale (MCS)		–	*0.39*	–
HADS				
A correlation of − 0.4 or greater between HADS anxiety scores		Convergent validity	–	*− 0.73*
A correlation of − 0.4 or greater between HADS depression scores		–	–	*− 0.67*
A correlation of − 0.4 or less between HADS anxiety scores		Discriminant validity	*− 0.40*	–
A correlation of − 0.4 or less between HADS depression scores		–	− 0.66	–

Analysed with Spearman Rho Correlation analysis. A priori corresponding constructs are marked in italics

P value for correlation coefficients < 0.001

timeframe of 14 days is often used in HRQoL studies [17, 31] which is largely accomplished in this study.

The relative reliability was good with $ICC_{2,1} > 0.80$ on each of the three HeartQoL scales [17]. The SDC values in our study indicate that a change greater than 0.5–0.7 points on respective scales would be required to be 95% certain that a score change would not be the result of measurement error, but rather the result of a real change. As HeartQoL scores range from 0 to 3 points, the SDC may be perceived as a relatively large measurement error. However, our results are consistent with those of Lee et al. [32].

The emotional subscale ceiling effect of 36% is above the recommended limit of 15% [17]. However, this is similar to the 27% ceiling effect observed in the original validation study [9] probably related to the fact that there are only four items in the emotional subscale. Although this indicates that further improvement in emotional health would not be detected by the HeartQoL instrument in 36% of the patient population, it may well be presumed that the respondents had reached a relatively stable emotional state 6–12 months post-surgery.

It should be noted that some items in the HeartQoL questionnaire do not coincide with the sternal precautions and specific activity restrictions after median sternotomy, thus the HeartQoL may be considered inappropriate if used during the first week's post-surgery.

Strengths of this study are that data are based on a nationwide survey, with a high response rate, low frequency of missing HeartQoL data, and further the study follows international methodological recommendations.

There are limitations to consider in this study. The risk of introducing bias is present as survey data are by nature subjective and self-reported information. A selection bias or non-respondent bias may also exist as there is an underrepresentation of woman, single/divorced or younger respondents. The relatively high level of missing data on the SF-36 may introduce bias into findings and in future validation studies a valid imputation method could be applied to avoid this. Using another HRQoL valve-specific instrument to test discriminant validity could have strengthened the study, but, to our knowledge, there is no existing validated heart valve-specific instrument.

Conclusion

The Danish HeartQoL questionnaire shows acceptable construct validity, internal consistency reliability, and test–retest reproducibility in patients following heart valve surgery in a cohort of Danish patients. Future studies should focus on assessing model structure (Mokken Scaling or confirmatory factor analysis) and the responsiveness of the HeartQoL questionnaire.

Table 6 Discriminative validity of the HeartQoL based on a priori hypotheses

Hypothesis	HeartQoL global	HeartQoL physical	HeartQoL emotional	Confirmed hypothesis
A				
Gender				Yes
Female	1.9 [1.4–2.5]	1.8 [1.1–2.5]	2.3 [1.8–3.0]	
Male	2.4 [1.8–2.9]	2.4 [1.6–2.8]	2.8 [2.0–3.0]	
P value	< 0.001	< 0.001	0.001	
B				
SF-36 health transition				Yes
Improved	2.4 [1.9–2.9]	2.4 [1.8–2.9]	2.8 [2.3–3.0]	
No change or deterioration	1.7 [1.1–2.5]	1.6 [0.9–2.5]	2.3 [1.5–3.0]	
P value	< 0.001	< 0.001	< 0.001	
C				
Charlson comorbidity Index				Yes
Scores ≤ 1	2.4 [1.7–2.8]	2.3 [1.5–2.8]	2.8 [2.0–3.0]	
Scores ≥ 2	1.9 [1.2–2.5]	1.8 [1.0–2.6]	2.5 [1.8–3.0]	
P value	< 0.001	< 0.001	0.002	
D				
HADS anxiety				Yes
Scores < 8	2.4 [1.9–2.8]	2.3 [1.6–2.8]	2.8 [2.3–3.0]	
Scores ≥ 8	1.1 [0.8–1.8]	1.2 [0.6–2.1]	1.3 [0.8–1.8]	
P value	< 0.000	< 0.001	< 0.001	
E				
HADS depression				Yes
Scores < 8	2.4 [1.9–2.8]	2.3 [1.6–2.8]	2.8 [2.3–3.0]	
Scores ≥ 8	1.1 [0.7–1.6]	0.9 [0.5–1.6]	1.3 [0.8–1.8]	
P value	< 0.001	< 0.001	< 0.001	
F				
Congestive heart failure				Yes
No	2.4 [1.6–2.8]	2.3 [1.4–2.8]	2.8 [2.0–3.0]	
Yes	2.0 [1.3–2.7]	1.9 [1.1–2.7]	2.3 [1.8–3.0]	
P value	0.006	0.005	0.014	

SF-36 short form 36 health survey, *HADS* the Hospital Anxiety and Depression Scale. HeartQoL values are reported in median and interquartile range [25th–75th quartile]. All analyses are assessed by Mann Whitney *U* test

Acknowledgements The authors thank the respondents for their participation and Jamal Uddin for assisting in data entry.

Funding The study was financially supported by The Danish Council for Strategic Research (10-092790), The Danish Heart Foundation, Edith and Olfert Dines Hansen Foundation (8868MT/IV), and the Research Council of the Region Zealand (12–000095/jun2014). The funding agencies had no influence on the interpretation of results.

Compliance with ethical standards

Conflict of interest Professor Neil Oldridge is a co-developer of the HeartQoL with no competing interest to report by the other authors.

Ethical approval The Danish Data Protection Agency approved the survey data (File Number: 2013-41-1643) and all local confidentiality and privacy requirements were met.

Informed consent Danish laws for register-based research do not require signed informed consent.

References

1. Supino, P. G., Borer, J. S., Preibisz, J., & Bornstein, A. (2006). The epidemiology of valvular heart disease: A growing public health problem. *Heart Failure Clinics, 2*(4), 379–393. https://doi.org/10.1016/j.hfc.2006.09.010.
2. Nkomo, V. T., Gardin, J. M., Skelton, T. N., Gottdiener, J. S., Scott, C. G., & Enriquez-Sarano, M. (2006). Burden of valvular heart diseases: A population-based study. *Lancet, 368*(9540), 1005–1011. https://doi.org/10.1016/s0140-6736(06)69208-8.
3. Blauwet, L. A., & Miller, F. A. Jr. (2014). Echocardiographic assessment of prosthetic heart valves. *Progress in Cardiovascular Diseases, 57*(1), 100–110. https://doi.org/10.1016/j.pcad.2014.05.001.

4. Berg, S. K., Zwisler, A. D., Pedersen, B. D., Haase, K., & Sibilitz, K. L. (2013). Patient experiences of recovery after heart valve replacement: Suffering weakness, struggling to resume normality. *BMC Nursing, 12*(1), 23. https://doi.org/10.1186/1472-6955-12-23.
5. Lapum, J., Angus, J. E., Peter, E., & Watt-Watson, J. (2011). Patients' discharge experiences: Returning home after open-heart surgery. *Heart and Lung, 40*(3), 226–235. https://doi.org/10.1016/j.hrtlng.2010.01.001.
6. Anker, S. D., Agewall, S., Borggrefe, M., Calvert, M., Jaime Caro, J., Cowie, M. R., et al. (2014). The importance of patient-reported outcomes: A call for their comprehensive integration in cardiovascular clinical trials. *European Heart Journal, 35*(30), 2001–2009. https://doi.org/10.1093/eurheartj/ehu205.
7. Rumsfeld, J. S., Alexander, K. P., Goff, D. C. Jr., Graham, M. M., Ho, P. M., Masoudi, F. A., et al. (2013). Cardiovascular health: The importance of measuring patient-reported health status: A scientific statement from the American Heart Association. *Circulation, 127*(22), 2233–2249. https://doi.org/10.1161/CIR.0b013e3182949a2e.
8. Oldridge, N., Hofer, S., McGee, H., Conroy, R., Doyle, F., & Saner, H. (2014). The HeartQoL: Part I. Development of a new core health-related quality of life questionnaire for patients with ischemic heart disease. *The European Journal of Preventive Cardiology, 21*(1), 90–97. https://doi.org/10.1177/2047487312450544.
9. Oldridge, N., Hofer, S., McGee, H., Conroy, R., Doyle, F., & Saner, H. (2014). The HeartQoL: Part II. Validation of a new core health-related quality of life questionnaire for patients with ischemic heart disease. *The European Journal of Preventive Cardiology, 21*(1), 98–106. https://doi.org/10.1177/2047487312450545.
10. Kristensen, M. S., Zwisler, A. D., Berg, S. K., Zangger, G., Gronset, C. N., Risom, S. S., et al. (2016). Validating the HeartQoL questionnaire in patients with atrial fibrillation. *The European Journal of Preventive Cardiology, 23*(14), 1496–1503. https://doi.org/10.1177/2047487316638485.
11. Zangger, G., Zwisler, A. D., Berg, K., Kristensen, S., Gronset, M. S., Uddin, C. N., J., et al (2018). Psychometric properties of HeartQoL, a core heart disease-specific health-related quality of life questionnaire, in Danish implantable cardioverter defibrillator recipients. *The European Journal of Preventive Cardiology, 25*(2), 142–149. https://doi.org/10.1177/2047487317733074.
12. De Smedt, D., Clays, E., Hofer, S., Oldridge, N., Kotseva, K., Maggioni, A. P., et al. (2016). Validity and reliability of the HeartQoL questionnaire in a large sample of stable coronary patients: The EUROASPIRE IV Study of the European Society of Cardiology. *The European Journal of Preventive Cardiology, 23*(7), 714–721. https://doi.org/10.1177/2047487315604837.
13. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *The Journal of Clinical Epidemiology, 63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006.
14. Lynge, E., Sandegaard, J. L., & Rebolj, M. (2011). The Danish National Patient Register. *Scandinavian Journal of Public Health, 39*(7 Suppl), 30–33. https://doi.org/10.1177/1403494811401482.
15. Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseasess, 40*(5), 373–383.
16. Sibilitz, K. L., Berg, S. K., Hansen, T. B., Risom, S. S., Rasmussen, T. B., Hassager, C., et al. (2013). Effect of comprehensive cardiac rehabilitation after heart valve surgery (CopenHeartVR): study protocol for a randomised clinical trial. *Trials, 14*, 104. https://doi.org/10.1186/1745-6215-14-104.
17. Vet, H. C. W. D. (2011). *Measurement in medicine: A practical guide (Practical guides to biostatistics and epidemiology)*. Cambridge: Cambridge University Press.
18. Ware, J. E. Jr. (2000). SF-36 health survey update. *Spine (Phila Pa 1976), 25*(24), 3130–3139.
19. Bjorner, J. B., Damsgaard, M. T., Watt, T., & Groenvold, M. (1998). Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. *The Journal of Clinical Epidemiology, 51*(11), 1001–1011.
20. Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research, 52*(2), 69–77.
21. Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*(6), 361–370.
22. Herrmann, C. (1997). International experiences with the Hospital Anxiety and Depression Scale–a review of validation data and clinical results. *Journal of Psychosomatic Research, 42*(1), 17–41.
23. Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using SPSS* (4ed.). Maidenhead: Open University Press.
24. Raykov, T., & Marcoulides, G. A. (2017). Thanks coefficient alpha, we still need you!. *Educational and Psychological Measurement*, https://doi.org/10.1177/0013164417725127.
25. Aguirre-Urreta, M. I., Ronkko, M., & McIntosh, C. N. (2018). A cautionary note on the finite sample behavior of maximal reliability. *Psychological Methods*. https://doi.org/10.1037/met0000176.
26. Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength and Conditioning Research, 19*(1), 231–240. https://doi.org/10.1519/15184.1.
27. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.
28. Hofer, S., Benzer, W., Schussler, G., von Steinbuchel, N., & Oldridge, N. B. (2003). Health-related quality of life in patients with coronary artery disease treated for angina: validity and reliability of German translations of two specific questionnaires. *Quality of Life Research, 12*(2), 199–212.
29. Leal, A., Paiva, C., Hofer, S., Amado, J., Gomes, L., & Oldridge, N. (2005). Evaluative and discriminative properties of the Portuguese MacNew Heart Disease Health-related Quality of Life Questionnaire. *Quality of Life Research, 14*(10), 2335–2341. https://doi.org/10.1007/s11136-005-7213-x.
30. Vandereyt, F., Dendale, P., Vanhees, L., Roosen, J., Hofer, S., & Oldridge, N. (2012). Psychometric properties of the Flemish version of the MacNew heart disease health-related quality of life questionnaire. *Acta Cardiologica, 67*(1), 31–39. https://doi.org/10.2143/ac.67.1.2146563.
31. Polit, D. F. (2014). Getting serious about test-retest reliability: A critique of retest research and some recommendations. *Quality of Life Research, 23*(6), 1713–1720. https://doi.org/10.1007/s11136-014-0632-9.
32. Lee, W. L., Chinna, K., Bulgiba, A., Abdullah, K. L., Abidin, I. Z., & Hofer, S. (2016). Test-retest reliability of HeartQoL and its comparability to the MacNew heart disease health-related quality of life questionnaire. *Quality of Life Research, 25*(2), 351–357. https://doi.org/10.1007/s11136-015-1097-1.