Psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q)

Middel, B.; Bouma, J.; de Jongste, M.J.L.; van Sonderen, F.L.P.; Niemeijer, M.; Crijns, H.; van den Heuvel, W.J.A.

Published in:
Clinical Rehabilitation

Publication date:
2001

Clinical Rehabilitation 2001; **15**: 489–500

Psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q)

Berrie Middel, Jelte Bouma Northern Centre for Healthcare Research (NCH), School of Medicine, University of Groningen, **Mike de Jongste** Department of Cardiology and Thoracic Surgery, University Hospital Groningen, **Eric van Sonderen** Northern Centre for Healthcare Research (NCH), School of Medicine, University of Groningen, **MG Niemeijer** Department of Cardiology, Martini Hospital Groningen, **Harry Crijns** Department of Cardiology and Thoracic Surgery, University Hospital Groningen and **Wim van den Heuvel** Northern Centre for Healthcare Research (NCH), School of Medicine, University of Groningen, Groningen, The Netherlands

Received 17th September 1999; returned for revisions 1st June 2000; revised manuscript accepted 1st August 2000.

Objective:To evaluate the psychometric properties of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q) in patients with atrial fibrillation.
Design: A prospective study of the patients who underwent DC electrical cardioversion.
Setting: Clinics of cardiology and thoracic surgery of the University Hospital in Groningen, the Netherlands.
Main outcome measures: The disease-specific MLHF-Q and generic measures of quality of life were administered. The sensitivity to change over time was tested with effect sizes (ES). Internal consistency of MLHF-Q scales was estimated with Cronbach's alpha. To evaluate the construct validity multitrait–multimethod analysis was applied. The 'known group validity' was evaluated by the comparison of mean scores and effect sizes between two groups of the New York Heart Association (NYHA) classification (NYHA I versus II–III). Stability of MLHF-Q scales was estimated in a subgroup of patients who remained stable. Perfect congruence analysis and factor analysis were applied to confirm the a priori determined structure.
Results: Cronbach's alpha was ≥0.80 of the MLHF-Q scales. Perfect congruence analysis (PCA) showed that the results resemble quite well the a priori assumed factor structure. Multitrait–multimethod analysis showed convergent validity coefficients ranging from 0.59 to 0.73 (physical impairment dimension) and 0.39 to 0.69 (emotional dimension). The magnitude of change can be interpreted as medium (ES = 0.50). The results of a 'test–retest' analysis in a stable group can be valued as satisfactory for the MLHF-Q scales (Pearson's $r > 0.60$). The physical dimension and the overall score of the MLHF-Q discriminated significantly between the NYHA I and II–III groups ($p < 0.001$) with large effect sizes (ES > 1.0).

Address for correspondence: Berrie Middel, Northern Centre for Healthcare Research, School of Medicine, University of Groningen, A Deusinglaan 1, 9713 AV Groningen, The Netherlands. e-mail: B.Middel@med.rug.nl

0269–2155(01)CR422OA

Conclusions: The MLHF-Q has solid psychometric properties and the outcome of the current study indicates that the MLHF-Q is an effective and efficient instrument.

Introduction

In assessing health-related quality of life and functional ability or health status, a distinction is made between disease-specific outcome measures developed to measure quality of life dimensions characteristic for patients having a particular disease, and generic instruments measuring more broadly defined dimensions of quality of life. Both types of instruments have their strengths and weaknesses.[1] An advantage of generic instruments is that they have a broad scope and can be used in many populations on a wide variety of diseases. A disadvantage is that general aspects of quality of life that are not significant for a specific disease will result in a less valid assessment of the concept of health-related quality of life in, for example, groups of (chronic) disease. Assessing only those aspects of quality of life which are determined to be due to a particular disease will result in a short instrument that will be more sensitive to detect change in disease-specific groups after (medical) interventions. A disadvantage of a disease-specific instrument is that study results are difficult to compare with other populations. In the current study, health-related quality of life was assessed with the disease-specific Minnesota Living with Heart Failure Questionnaire (MLHF-Q), the generic RAND-36 or Short Form (SF-36),[2] the Multidimensional Fatigue Inventory (MFI-20)[3–5] and the Hospital Anxiety and Depression Scale (HADS).[6,7] The data were appropriate to conduct a validation study to estimate the sensitivity to change (responsiveness), the reliability and validity of the disease-specific MLHF-Q to obtain data for its future use in (Dutch) clinical evaluation studies.

The MLHF-Q consists of 21 items and addresses a wide range of health-related quality of life aspects.[8] In this article the psychometric properties of the Dutch version of the MLHF-Q scales are evaluated and validated with conceptually similar dimensions of generic instruments, the RAND-36, the HADS and the MFI-20.

All instruments are self-report measures of quality of life on the dimensions of physical, mental or social well-being. The psychometric properties of the MLHF questionnaire have been evaluated previously in its English version and the instrument has been used as an outcome measure in clinical trials in the context of the American health care system.[9–14] In other countries there have been only a few studies on the evaluation of the reliability and validity of the MLHF-Q up till now.[15,16] The RAND-36 was chosen as the generic counterpart because it is a generally accepted and well-validated instrument, it is a short questionnaire with known psychometric properties,[17–20] it resembles closely the MLHF-Q dimensions and it is available in a Dutch version.[21,22]

The objectives of this study were:

- To compare the results from the MLHF-Q with the RAND-36, HADS and MFI-20 in terms of reliability and sensitivity to detect change over time. It was hypothesized that the MLHF-Q would demonstrate a comparable magnitude of change over time.
- To compare the results of the questionnaire's clinical validity. It was hypothesized that the MLHF-Q would demonstrate that the more severely angina was rated by the New York Heart Association (NYHA) classification, the greater the deterioration in the patient's quality of life. The change assessed in a group of patients who remained clinically unchanged or stable was hypothesized to be due to chance fluctuation.
- To find support for the factor structure originally found by Rector and Cohn[12] in our data.
- To provide empirical evidence that the MLHF-Q scale measures the underlying constructs of physical and emotional impairments it is reputed to represent.

The purpose of the present study was to use the data of a treatment–outcome study to determine the performance of the MLHF-Q. The results of the clinical efficacy study will be published elsewhere.

Methods

Consecutive patients scheduled for DC electrical cardioversion were included in this study. Patients presented with atrial fibrillation and atrial flutter and were treated at the department of cardiology and thoracic surgery of the University Hospital in Groningen.

Out of the 60 consecutive candidates for DC electrical cardioversion screened for inclusion, five patients died within 12 months after the completion of the first questionnaire. One year after the first visit to the clinic, 44 patients out of 55 (80.0%) returned the questionnaire used for analysis of reliability and validity of the MLHF-Q.

All patients completed the questionnaires as a baseline assessment before the first treatment (DC electrical cardioversion) in the department of cardiology and thoracic surgery of the University Hospital Groningen. The patients were invited to participate in the study by the cardiologist and after informed consent the patients completed the questionnaires, undisturbed, in a separate room. The cardiologist was blinded to the information of the questionnaires. The second and third assessment was at home, three and 12 months after the first electrical cardioversion respectively. The questionnaires were returned in a pre-paid envelope to the Northern Centre for Healthcare Research of the University of Groningen.

Measurements

Both demographic characteristics of the patients and relevant medical background variables were administered with standard or usual questions and items in the medical examination procedure at the first visit to the outpatient clinic. To assess the impact of the treatment on daily physical, emotional and social functioning four instruments were used. The RAND-36 is a generic instrument and consists of 36 items that contribute to eight scales that measure the following aspects of health: 'physical functioning' (10 items), 'social functioning' (2), 'role limitations due to physical problems' (4), 'role limitations due to emotional problems' (3), 'mental health' (5), 'energy/vitality' (4), 'pain' (2) and 'general health perception' (5). The one-item scale on change in perceived health was not used in the transformation of scores into a scale because the MLHF questionnaire does not contain an item assessing change in perceived health. The RAND-36 item scores are summed and transformed to eight scales, each with scores between 0 and 100, where 0 represents the unhealthiest state and 100 the best health state possible.[2,22]

The MLHF-Q is a disease-specific instrument composed of 21 items and three scales that measure: the physical dimension (8 items), the emotional dimension (5 items) and the overall score on health-related quality of life (21 items). Eight separate items, which do not assess a single construct or dimension of health-related quality of life, measure social and economical impairments patients relate to their heart failure and are part of the overall score.The total score has a range between 0 and 105, the physical dimension (subscale) between 0 and 40, the emotional dimension (subscale) between 0 and 25 and the separate items on the socio-economic impairments between 0 and 40.

High scores on the MLHF-Q scales indicate a high negative impact of heart disease on the assessed aspects of quality of life.

The Multidimensional Fatigue Inventory (MFI-20) consists of 20 items and five subscales (general, physical, activity, motivation and cognition). Each scale consists of four items and has a range from 4 to 20 and its total score ranges from 20 to 100. High scores indicate high fatigue. The subscales anxiety and depression of the Hospital Anxiety and Depression Scale (HADS) have a range between 0 and 21. A score of 7 or lower identifies 'non cases', 8 to 10 'doubtful cases' and a score ≥11 'definite cases'.

Quantitative analysis

The features of the distribution of scores on the conceptually similar dimensions of the MLHF-Q, MFI-20, HADS and RAND-36 were computed. Mean scores, standard deviations and the percentage of patients with the maximum possible score (ceiling) and the minimum possible score (floor) are represented.

In the examination of the construct validity of the MLHF-Q, scales of all instruments were used in the analysis. It was hypothesized that the scales

that are strongly conceptually associated would show strong correlations and scales that are conceptually associated more weakly would demonstrate lower correlation coefficients.

In this study the internal consistency of the MLHF-Q, RAND-36, HADS and MFI-20 scales was tested with Cronbach's α[23] to make comparisons between the instruments' mean alphas. An α coefficient >0.80 was considered as sufficient,[24] irrespective of the number of items. Perfect congruence analysis and factor analysis were applied to confirm the a priori determined structure on which Rector and Cohn[12] have selected the items.

Test–retest stability of the MLHF-Q scales was assessed with correlation coefficients between baseline and three months after cardioversion in a group in which the treatment was not successful (that showed no sinus rhythm three months after the first electrical cardioversion), so their health status remained unchanged or stable. Although the test–retest procedure was not carried out by sending the questionnaire shortly after the first completion, we were interested in the variability of the MLHF-Q scores between two points in time within a group whose condition remained stable. However, high test–retest correlation coefficients as such do not give us information about the changes in time between baseline and three-month outcome scores and therefore we tested the hypothesis that the change over time in a stable group is due to chance fluctuations. The Wilcoxon matched-pairs signed-rank test was used due to the non-normal distribution of the outcome assessments.

To estimate the responsiveness, the ability of an instrument to detect the magnitude of change over time within one group, we used Cohen's effect size statistic *d* for paired observations.[25] As the variance of the post-test measure is partly explained by the pretest scores, estimating the magnitude of the change between baseline and post-test in the treated group requires adjustment of the effect size *d'* for the correlation (*r*) between the baseline and post-test scores[26,27]:

$$d = \frac{d'}{\sqrt{1 - r}} \qquad d' = \frac{\bar{x}_{\text{baseline}} - \bar{x}_{\text{post-test}}}{\text{SD}\ (x_{\text{baseline}}\ x_{\text{post-test}})}$$

where *d'* is the effect size = mean change/pooled SD baseline and post-test score; *d* is the effect size adjusted for *r*; *r* is the correlation coefficient between repeated measurements.

An effect size of 0.20 has to be interpreted as a small effect, an effect size of 0.50 a medium effect and an effect size of >0.80 a large effect.[25,28]

To evaluate the ability of the MLHF-Q to discriminate between subgroups of patients which are known to differ on an accepted classification of the seriousness of the disease, the 'known groups validity' of the MLHF-Q scales was tested.[29] The Mann–Whitney *U* Wilcoxon rank sum test was used because of the non-normal distribution of the variables in the analysis. The grouping condition was NYHA classification I versus II and III (due to the small number of observations class II and III were combined).[30] Cohen's effect size *d'* for unrelated samples to estimate the magnitude of the difference in mean scores between these groups was calculated by dividing the mean difference score by the pooled standard deviation for groups with unequal number of observations[31]:

$$d' = \frac{\bar{x}_{\text{NYHA I}} - \bar{x}_{\text{NYHA II-III}}}{\text{SD}\ (x_{\text{NYHA I}}\ x_{\text{NYHA II-III}})}$$

Results

Table 1 shows the descriptive statistics of the sample. The mean (range) age of the patients in the study was 61.5 (range 28–87) years. A minority of the patients were female (35%). The majority of the patients had one or more heart diseases or other relevant diseases in addition to atrial fibrillation (AF). Only six persons had AF without any other disease. Almost half of the patients (46.7%) had two or more diseases next to AF. A relatively large group (41.7%) was treated for the first time for AF. The mean score on the NYHA classification (range 1–4) of 1.9 indicates a moderate severity of the underlying disease.

Distribution of scores, internal consistency and responsiveness

Mean baseline and post-test (1 year) scores, standard deviations and the percentages of

patients with the maximum and minimum scores, are shown in Table 2. A study of the distribution of scores of the MLHF-Q scales showed a skewness in the direction of positive functioning or little or no impairment. The RAND-36 data showed the same tendency for four scales (social functioning, emotional role functioning, pain and health perception). The RAND-36 scale 'physical role functioning' showed a tendency towards the opposite direction. Three conceptually related scales of the MFI ('physical', 'activity' and 'general' feelings of fatigue) were skewed in the direction of little impact on health-related quality of life, while the cognition scale was skewed in the negative direction. The Cronbach's alphas, the internal consistency coefficients, of the MLHF-Q, RAND-36, HADS and MFI-20 scales are also shown in Table 2. The internal consistency of the MLHF-Q scales had a satisfactory level of reliability ($\alpha > 0.80$).[24] Only the RAND-36 scales 'social functioning' and 'general health perception' and the MFI-scale 'cognition' were below this level (0.79, 0.79 and 0.76 respectively). The reliability coefficients of the MLHF-Q scales remained satisfactory one year after enrolment.

Table 1 Patient characteristics at study enrolment ($n = 60$)

	No. (%)
Gender	
Men	39 (65.0)
Women	21 (35.0)
Mean age (years)	61.5 (SD 12.7)
Marital status	
Married/living with partner	39 (60.0)
Widowed/unmarried/divorced	16 (24.6)
Missing value	5 (15.4)
Disease	
AVD	12 (20.0)
MVD	12 (20.0)
Hypertension	16 (26.7)
Congenital heart disease	7 (11.7)
Coronary artery disease	11 (18.3)
Cardiomyopathy	4 (6.7)
Hyperthyroidism	2 (3.3)
Chronic obstructive pulmonary disease	9 (15.0)
Miscellaneous	16 (26.7)
No disease	6 (10.0)
1 disease	26 (43.3)
2 diseases	21 (35.0)
3–4 diseases	7 (11.7)
Mean NYHA classification	1.9 (SD 0.6)

AVD, aortic valve disease; MVD, mitral valve disease.

The scales of the MLHF-Q at baseline assessment yielded internal consistency estimates (mean $\alpha = 0.85$; range = 0.82–0.88) equal to the RAND-36 (mean $\alpha = 0.86$; range = 0.79–0.93) and somewhat higher than those of the HADS (mean $\alpha = 0.83$; range = 0.83–0.84) and MFI-20 (mean $\alpha = 0.84$; range = 0.76–0.88). The ability to detect change over time within one group with paired observations was estimated with the effect size proposed by Cohen.[25] An effect size of 0.20 has to be interpreted as a small effect, an effect size of 0.50 a medium effect and an effect size of >0.80 a large effect. Large effect sizes were not found. The MLHF-Q scales showed medium effect sizes. The RAND-36 scales 'role limitations due to emotional problems' and 'pain' demonstrated small effect sizes and 'general health perception' showed no ability to detect change between baseline and one year outcome assessment. The HADS-anxiety scale and the MFI-20 general fatigue scale showed medium effect sizes. The physical and emotional dimensions of the RAND-36, HADS and MFI-20 demonstrate comparable indicators of change over time within this particular group.

Item analysis

The MLHF-Q contains three dimensions or scales: a physical dimension, an emotional dimension and a global quality of life dimension. A comparison was made with the results of the factor analysis of Rector and Cohn.[2] Their data provided us with an a priori assumed four-factor structure that was forced in order to evaluate the congruence of our data with the original structure. Therefore, a computer program for simultaneous component analysis (SCA) for variables measured in two or more populations was applied.[32] The four a priori assumed factors based on the structure in the data of Rector and Cohn explained 58% of the total variance as a result of the SCA-perfect congruence analysis (PECON).[33] A principal component analysis with rotation according to the varimax criterion was performed without the constraints of the structure elaborated by Rector and Cohn. In this

Table 2 Means, standard deviations, minimum and maximum scale scores, Crohnbach's alphas, Pearson's correlations $t_1 - t_{12}$ and within-group effect size for paired observations ($n = 44$)

	Pretest					Post-test						
Dimension	Mean	SD	% floor	% ceiling	Reliability	Mean	SD	% floor	% ceiling	Reliability	Effect size[a]	Pearson's *r*
MLHF-Q												
Physical dimension (0–40)	14.2	9.6	15.9	2.3	0.88	10.4	10.3	15.9	2.3	0.91	0.65	0.67
Emotional dimension (0–25)	5.9	5.7	20.5	2.3	0.82	3.8	4.5	31.8	2.3	0.81	0.56	0.51
Overall score (0–105)	28.5	19.6	11.4	2.3	0.91	21.6	20.8	13.6	2.3	0.94	0.59	0.67
RAND-36												
Physical functioning (0–100)	56.6	28.7	2.3	6.8	0.93 (0.92)[b]	66.1	27.1	11.4	2.3	0.93	0.63	0.71
Social functioning (0–100)	60.4	25.3	13.6	4.7	0.79 (0.71)	72.3	25.2	31.8	2.3	0.78	0.72	0.56
Role–physical (0–100)	27.3	39.3	15.9	56.8	0.91 (0.90)	51.7	45.1	38.6	34.1	0.91	0.77	0.46
Role–emotional (0–100)	54.8	44.7	31.8	40.9	0.90 (0.86)	62.6	44.3	54.5	25.0	0.90	0.25	0.52
Pain (0–100)	80.9	23.6	47.7	2.3	0.90 (0.93)	82.6	21.9	52.3	2.3	0.89	0.13	0.67
Mental health (0–100)	64.6	21.7	6.8	4.7	0.83 (0.85)	72.5	17.8	2.3	2.3	0.84	0.54	0.47
Energy/vitality (0–100)	48.2	24.2	2.3	4.9	0.86 (0.82)	58.6	22.4	2.3	2.3	0.84	0.65	0.53
Health perception (0–100)	56.0	21.2	6.8	2.3	0.79 (0.82)	55.0	20.0	2.3	4.5	0.76	0.07	0.51
HADS												
Anxiety	5.7	3.9	4.5	2.3	0.83	4.2	3.3	15.9	2.3	0.81	0.61	0.53
Depression	6.0	4.6	6.8	4.5	0.84	5.3	4.5	11.4	2.3	0.86	0.26	0.74
MFI-20												
General	12.6	5.2	6.8	11.4	0.87	10.8	5.3	11.4	9.1	0.89	0.58	0.66
Physical	12.1	4.5	7.0	11.4	0.85	10.9	5.1	9.1	4.5	0.90	0.37	0.49
Activity	12.3	5.1	4.5	9.1	0.88	10.9	5.3	13.6	6.8	0.90	0.41	0.54
Motivation	10.7	4.7	4.5	4.5	0.76	10.2	4.5	15.9	2.3	0.80	0.20	0.67
Cognitive	7.4	3.5	34.1	2.3	0.82	8.0	3.9	27.3	2.3	0.86	0.24	0.52
Overall fatigue	54.1	18.3	2.4	2.4	0.94	49.5	20.2	2.3	2.3	0.95	0.40	0.65

[a]Effect size *d* for paired observations.[26]
[b]Reliabilities of a general Dutch municipality population.[21,22]
MLHF-Q, Minnesota Living with Heart Failure Questionnaire; SF-36, Short Form 36; HADS, Hospital Anxiety and Depression Scale; MFI-20, Multidimensional Fatigue Inventory.

analysis the four factors explained 61% of the total variance. This difference of 3% indicates an acceptable discrepancy, but still indicates an insufficient recognition in our data. A fourth socio-economic dimension of impairments that patients relate to their heart failure was suggested by Rector *et al*.[9] but in the current study the items did not load on a socio-economic component. As demonstrated by the matrix (Table 3), 6 out of the 21 items had very high loadings (>0.70) and 13 items had high loadings (>0.50–<0.70) on their respective factors. Only one item had a high loading on two factors (impairment because of ankle oedema). On face value we may conclude that the results closely resemble the findings of Rector and Cohn.[12] A closer inspection of the four-factor solution, however, shows some deviations from the original factor structure: two items of the physical dimension identified by Rector and Cohn ('making sleeping well at night difficult' and 'making your relating to or doing things with your friends or family difficult') have a high loading on factors 3 and 4, representing the impairments on a heterogeneous set of health-related aspects of heart failure. Factor 2 demonstrates high loadings of the items on the physical dimension. Although all the items of the emotional dimension had high loadings on factor 1, the following items also showed high loadings: 'making going away from home difficult' (physical dimension), 'ankle oedema', 'hospitalization' and 'medical costs' (socio-economic impairments) on this factor.

Construct validity

In this study we attempted to provide evidence that the MLHF-Q scales measure the underlying constructs of physical and emotional impairments they are reputed to represent.

The multitrait–multimethod approach outlined by Campbell and Fiske[34] was used to assess the convergent and divergent validity of the MLHF-Q measures of physical and emotional impairment.

Convergent validity (i.e. evidence that we are measuring what we purport to measure) is provided by data that show that different measures of conceptually related dimensions of health-

Table 3 Principal components factor analysis with varimax rotation of the MLHF-Q

	Factor 1	Factor 2	Factor 3	Factor 4
Making you stay in a hospital	0.69[a]	0.23	0.10	–0.13
Making you feel you are a burden to your family	0.62[b]	0.12	0.36	–0.34
Making you feel depressed	0.70[b]	0.06	0.33	0.26
Making you worry	0.69[b]	0.16	–0.05	0.14
Making you feel a loss of self-control in your life	0.64[b]	0.37	0.17	0.09
Making going away places away from home difficult	0.67[c]	0.35	0.11	0.21
Making it difficult for you to concentrate or remember things	0.65[b]	0.01	0.01	0.33
Costing you money for medical care	0.47[a]	0.18	0.25	–0.29
Causing swelling your ankles, legs, etc.	0.52[a]	0.52	–0.01	–0.02
Making walking about or climbing stairs difficult	0.26	0.83[c]	0.10	0.02
Making working around the house or yard difficult	0.31	0.79[c]	0.26	0.19
Making you sit or lie down to rest during the day	0.40	0.73[c]	0.11	–0.01
Making you tired, fatigued or low on energy	0.27	0.54[c]	0.31	0.25
Making you short of breath	–0.06	0.60[c]	0.31	0.26
Making sexual activities difficult	0.02	0.13	0.82[a]	0.16
Making you eat less of the foods you like	0.06	–0.07	0.76[a]	0.16
Making recreational pastimes, sports or hobbies difficult	0.13	0.49	0.68[a]	0.17
Making your relating to or doing things with your friends or family difficult	0.28	0.26	0.47[c]	0.24
Side effects from medications	0.22	0.32	0.56[a]	–0.06
Making working to earn a living difficult	0.11	0.10	0.21	0.73[a]
Making sleeping well at night difficult	0.10	0.22	0.39	0.60[c]

[a]Single items used in the construction of the overall score.
[b]Emotional dimension.
[c]Physical dimension.

Table 4 Multitrait–multimethod matrix for the emotional and mental dimensions of health-related quality of life ($n = 60$)

Constructs	1	2	3	4	5	6	7	8	9	10
1. MLHF physical dimension										
2. RAND-36 physical functioning	–0.65									
3. RAND-36 role – physical	–0.60	**0.64**								
4. RAND-36 energy/vitality	–0.73	**0.66**	**0.61**							
5. MFI-20 physical	0.63	–0.72	–0.59	–0.70						
6. MLHF emotional dimension	**0.44**	–0.28	–0.35	–0.34	0.37					
7. RAND-36 mental health	0.32	**0.35**	**0.31**	**0.53**	–0.35	–0.46				
8. RAND-36 role – emotional	0.33	**0.27**	**0.49**	**0.43**	–0.27	–0.47	**0.50**			
9. HADS anxiety	0.30	–0.24	–0.30	–0.25	0.42	0.43	–0.60	–0.59		
10. HADS depression	0.44	–0.46	–0.37	–0.46	0.35	0.44	–0.59	–0.39	**0.55**	

All correlations $p < 0.01$; corresponding dimensions are printed bold.
The areas surrounded by solid lines are the heterotrait–heteromethod triangles.
The area surrounded by broken lines comprises the coefficients for variables that have no trait in common (the hetereotrait–monomethod coefficients are depicted bold in both areas).

related quality of life are highly correlated.[35,36] In addition, we expect that each of the measures of physical and emotional dimensions of quality of life measures a different construct (i.e. that the 'physical fatigue' scale does not measure depression (divergent validity).

In Table 4 multitrait–multimethod matrices were constructed for each of the assessed dimensions of quality of life (physical and emotional). Evidence of convergent validity is drawn from examination of the coefficients in the heterotrait–heteromethod triangles, enclosed by solid lines in Table 4. We also expect some association between the scales measuring dimensions of, for example, physical quality of life if the same questionnaire (method) was used and items were not presented in a randomized order (correlated measurement error). These heterotrait–monomethod coefficients are depicted in bold. In the area enclosed by broken lines the coefficients between variables that have no trait in common are shown.

The correlations between the three generic methods and the MLHF-Q scale assessing physical impact on quality of life are, as expected, high and have the same magnitude as the heterotrait–monomethod coefficients. The correlations between the three generic methods and the MLHF-Q scale assessing emotional impact, while statistically significant, are moderate (except the correlation between the HADS depression scale with the RAND 'role emotional' scale).

To demonstrate divergent validity the multitrait–monomethod correlation coefficients must be higher than the correlation coefficients for variables that have neither trait nor method in common. The values that represent relations between the components of physical and emotionally impaired quality of life, which are represented in the area enclosed by broken lines, are of interest. Most of the scales that are supposed to measure different constructs are weakly correlated, independent from the method that is used. In accordance with our expectation, some correlations were of moderate magnitude simply due to shared method variance (printed in bold).

This analysis provides reassurance that with the MLHF-Q we are measuring physical impairment and that there is convergence among methods. The emotional impairment component, however, is moderately associated with the other methods.

Test–retest

If a quality of life instrument like the MLHF-Q is developed to be used as an evaluative instrument in clinical trials, one of the conditions which should be fulfilled is that it has the ability to demonstrate stability over time in subjects whose health status does not change (test–retest reliability).[36]

Table 5 shows the test–retest correlation coefficients after a period of three months of stability in health status without serious cardiac events.

Table 5 Means, standard deviations (SD), test–retest correlations (*r*) and difference in scores between baseline and three-month outcome in a group of patients that did not show improvement in sinus rhythm

	Mean (*t*1)	SD (*t*1)	Mean (*t*2)	SD (*t*2)	*r*	*z*-score	*p*-value
MLHF-Q							
Physical dimension (*n* = 19)	14.39	9.59	15.91	12.91	0.70	–0.23	0.82
Emotional dimension (*n* = 19)	6.05	5.77	4.63	4.69	0.63	–1.47	0.14
Overall score (*n* = 19)	29.79	18.65	26.00	20.34	0.73	–1.28	0.20

The results can be valued as satisfactory for all MLHF-Q scales. However, although we can interpret the test–retest correlation coefficients as satisfactory, these estimates of linear relationships do not provide information about the existence of significant change in a selected group of stable patients. To test the statistical significance of the change between baseline and three-month outcome the Wilcoxon matched-pairs signed-rank test was used because of the non-normal distribution of the MLHF-Q scales. None of the MLHF-Q scales demonstrated significant change.

Known group validity

To evaluate the ability of the MLHF-Q dimensions to discriminate between so-called 'known groups' who should differ based on the cardiologists' (blinded) classification of the severity of the disease, the study sample was divided into two subgroups: NYHA classification I and II–III. The results of the analysis of the ability of the MLHF-Q scales to discriminate between 'known groups' are presented in Table 6. The physical dimension and the overall score of the MLHF-Q discriminated sharply between the NYHA II–III and I groups ($p < 0.001$) with large effect sizes. The MLHF emotional dimension also discriminated clearly between these groups ($p = 0.01$) but with a moderate effect size.

Discussion

To what extent does the Dutch version of the MLHF-Q measure the desired underlying concept or reflect what it is supposed to measure? In this study, the MLHF-Q construct validity was determined by higher and significant correlation coefficients between the MLHF-Q scales and corresponding dimensions of the MFI-20, HADS and the RAND-36 and by lower correlations with noncorresponding dimensions of health-related quality of life of these instruments. The MLHF-Q 'physical' dimension showed higher correlations with the RAND-36 scales 'social functioning', 'energy vitality', 'health perception' and 'pain', indicating that these domains of quality of life, which are not tagged by the MLHF-Q, are more likely to be associated with physical limitations in this study group.

One of the great advantages in clinical trials is

Table 6 Discriminative ability of the MLHF-Q between NYHA classification groups

	NYHA class I (*n* = 15)		NYHA class II and III (*n* = 38)		Rank sum[a]		
Dimension	Mean	SD	Mean	SD	*z*-value	*p*-value	Effect size[b]
MLHF-Q							
Physical dimension (0–40)	5.5	7.1	16.7	9.0	–3.8	0.0001	1.31
Emotional dimension (0–25)	3.5	6.4	6.5	5.3	–2.5	0.01	0.53
Overall score (0–105)	11.7	16.5	33.9	18.1	–3.6	0.0003	1.25

[a]Mann–Whitney *U*, one-sided.
[b]Estimation of the effect size used Cohen's *d* for independent samples when $n_1 \neq n_2$, which is defined as the difference in mean scores divided by the pooled standard deviation: est. $\sigma = \sqrt{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2 / (N_1 - 1) + (N_2 - 1)}$.

Clinical messages

- Patients considered for cardioversion of atrial fibrillation were studied to validate the Dutch version of the Minnesota Living with Heart Failure Questionnaire (MLHF-Q), by investigating responsiveness, reliability, validity and effect size.
- Outcomes showed that the MLHF-Q is an effective and efficient instrument to assess clinically important changes in health-related quality of life.

that the MLHF-Q is short; but its disadvantage is that it does not cover other relevant domains of quality of life impairment, such as impairment of social functioning or vitality. In the detection of change over time (pre- and post-test) the MLHF-Q performs equally well compared with the RAND-36, estimating the same standardized mean change score expressed in effect sizes that are interpreted as medium effect for both instruments on physical and emotional functioning. We hypothesized a greater responsiveness because the MLHF-Q should have greater precision due to the disease-specific operationalized items of the domains physical and emotional impairment. An alternative explanation for the failure to detect greater changes may be related to the selected group of patients: first, the questionnaire was developed to assess health-related quality of life associated with heart failure, which does not exist in every subject within this group. Secondly, disappearance of the atrial fibrillation (AF) probably has no strong impact on health-related quality of life because of the fact that in 90% of the subjects the underlying diseases in addition to AF still exist. However, it is to be expected that in 'before–after' intervention studies the MLHF-Q will show the ability to detect the appropriate magnitude of change over time. This expectation is confirmed by the results of our study, which showed that the MLHF-Q is sensitive to change within and between groups, even if the differences are small.

In the ability to discriminate between 'known groups' the magnitude of the difference on the physical dimension of the MLHF-Q was large (effect size > 1) and statistically significant ($p < 0.001$). The emotional impact on quality of life showed a statistically significant difference (accompanied with a moderate effect size) between NYHA I and II–III classified subjects. The substantial difference between both estimates of the magnitude of the difference between NYHA I and II–III may be determined by the dominant physical component of the NYHA classification. The correlations with conceptually related emotional dimensions (scales) were, while significant, of moderate magnitude. The cultural differences between the American and Dutch society, in combination with semantic differences in the translation of the items, are probably explanatory factors. 'Making your going places away from home difficult' and 'making you stay in a hospital' are probably more associated with the emotional impact of disturbing the relationship with significant others than with physical inhibition in the Dutch translation. The results of the current study indicate that the application of the MLHF-Q will enable Dutch researchers to assess health-related quality of life in clinical trials in which clinically relevant change will occur.

Acknowledgement

We would like to thank Dr Thomas S Rector for giving us permission to use the Minnesota Living with Heart Failure Questionnaire for scientific purposes.

References

1 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989; **27S**: S217–S232.
2 International Resource Center for Health Care Assessment (IRC). *How to score the MOS 36-item Short Form Health Survey (SF36): SF-36™ scoring rules*. Boston: New England Medical Center Hospitals, 1991.
3 Smets EMA, Garssen B, Bonke B. *Het Meten van vermoeidheid met de Multidimensionele Vermoeidheids Index (MVI-20)/Multidimensional Fatigue Inventory*. Een Handleiding. Amsterdam: Afdeling Medische Psychologie AMC, Universiteit van Amsterdam, 1995.
4 Smets EMA, Garssen B, Bonke B *et al*. The Multidimensional Fatigue Inventory (MFI): Psychometric qualities of an instrument to assess

fatigue. *J Psychosom Res* 1995; **39**: 315–25.
5 Smets EMA, Garssen B, Cull A *et al*. The application of the Multidimensional Fatigue Inventory (MFI-20) in cancer patients receiving radiotherapy. *Br J Cancer* 1996; **73**(2): 241–45.
6 Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; **67**: 361–70.
7 Spinhoven PH, Ormel J, Sloekers PPA *et al*. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychol Med* 1997; **27**(2): 363–70.
8 Guyatt GH. Measurement of health-related quality of life in heart failure. *J Am Coll Cardiol* 1993; **22**(4 (suppl A)): 185A–91A.
9 Rector TS, Kubo SH, Cohn JN. Patients' self-assessment of their congestive heart failure. Part 2: content, reliability and validity of a new measure, The Minnesota Living with Heart Failure Questionnaire. *Heart Failure* 1987; **3**: 198–209.
10 Kubo SH, Gollub S, Bourge R *et al*. Beneficial effects of Pimobendan on exercise tolerance and quality of life in patients with heart failure. *Circulation* 1992; **85**(3): 942–49.
11 Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure Questionnaire as a measure of therapeutic response to Enalapril or placebo. *Am J Cardiol* 1993; 71(May 1): 1106–107.
12 Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota Living with Heart Failure questionnaire: Reliability and validity during a randomized, double blind, placebo-controlled trial of pimobendan. *Am Heart J* 1992; October, 124(4): 1017–25.
13 Colucci WS, Packer M, Bristow MR *et al*. Carvedilol inhibits clinical progression in patients with mild symptoms of heart failure. US Carvedilol Heart Failure Study Group. *Circulation* 1996; **94**(11): 2800–806.
14 Bulpitt CJ. Quality of life with ACE inhibitors in chronic heart failure. *J Cardiovasc Pharmacol* 1996; **27**(suppl 2): S31–S35.
15 Cohen-Solal A, Caviezel B, Laperche T *et al*. Analyse critique des échelles de qualité de vie en cardiologie; applications à l'ínsuffisance cardiaque. *Arch Mal Coeur* 1994; **87**(IV): 71–77.
16 Metra M, Nodari S, Garbellini M *et al*. The effects of mid- and long-term administration (3-4 years) of carvedilol in patients with idiopathic dilated cardiomyopathy. *Cardiologia* 1997; **42**(5): 503–12.
17 Jenkinson C, Layte R, Lawrence K. Development and testing of the Medical Outcomes Study 36-Item Short Form Health Survey summary scale scores in the United Kingdom. Results from a large-scale survey and a clinical trial. *Med Care* 1997; **35**(4): 410–16.
18 McHorney CA, Ware JEJ, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; **31**(3): 247–63.
19 Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993; **306**(6890): 1437–40.
20 Van der Zee K, Sanderman R, Heyink JW *et al*. Psychometric qualities of the RAND 36-item Health Survey 1.0: a multidimensional measure of general health status. *Int J Behav Med* 1996; **3**: 104–22.
21 Van der Zee K, Sanderman R, Heyink JW. De psychometrische kwaliteiten van de MOS Short Form health Survey (SF-36) in een Nederlandse populatie. [The psychometric properties of the SF-36 in a Dutch population]. *Tijdschrift Soc Gezondheidszorg* 1993; **71**: 183–91.
22 Van der Zee K, Sanderman R. *Het meten van de algemene gezondheidstoestand met de RAND-36, een handleiding*. [The assessment of general health status with the RAND-36(manual)]. Noordelijk Centrum voor Gezondheidsvraagstukken, 1993.
23 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrica* 1951; **16**: 297–334.
24 Nunnally JC. *Psychometric theory*, second edition. New York: McGraw-Hill, 1978.
25 Cohen J. *Statistical power analysis for the behavioural sciences*, revised edition. New York: Academic Press, 1977.
26 Middel B, Kuipers-Upmeijer H, Bouma J *et al*. Effect of intrathecal baclofen delivered by an implanted programmable pump on health-related quality of life in patients with severe spasticity. *J Neurol Neurosurg Psychiatry* 1997; **63**: 204–209.
27 Vulink NCC, Overgaauw DM, Jessurun GA *et al*. The effects of spinal cord stimulation on quality of life in patients with therapeutically chronic refractory angina pectoris. *Neuromodulation* 1999; **2**: 33–40.
28 Cohen J. A power primer. *Psychol Bull* 1992; **112**(1): 155–59.
29 Crocker L, Algina J. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston, 1986.
30 New York Heart Association. *Diseases of the heart and blood vessels: nomenclature and criteria for diagnosis*. Boston: Little Brown, 1964.
31 Lipsey MW. *Design sensitivity. Statistical power for experimental research*. London: SAGE Publications, 1990.
32 Kiers HAL. *SCA, a program for simultaneous components analysis of variables measured in two or more populations (Manual)*. ProGAMMA, University of Groningen, 1990.
33 Ten Berge JMF. Rotation to perfect congruence and cross-validation of component weights across populations. *Multivariate Behav Res* 1986; **21**: 262–66.

34 Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychol Bull* 1959; **56**: 81–105.
35 Teresi JA, Golden RR, Gurland BJ, Wilder DE, Benett RG. Construct validity of indicator-scales developed from the comprehensive assessment and referral evaluation interview schedule. *J Gerontol* 1984; **39**(2): 147–57.
36 Varni JW, Seid M, Rode CA. The PedQL™: Measurement model for the pediatric quality of life inventory. *Med Care* 1999; **37**(2): 126–39.
37 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Disabil* 1985; **38**(1): 27–36.