

Development and Evaluation of the Kansas City Cardiomyopathy Questionnaire: A New Health Status Measure for Heart Failure

C. Patrick Green, MD,* Charles B. Porter, MD, FACC,*† Dennis R. Bresnahan, MD, FACC,*† John A. Spertus, MD, MPH, FACC*†

Kansas City, Missouri

-
- OBJECTIVES** To create a valid, sensitive, disease-specific health status measure for patients with congestive heart failure (CHF).
- BACKGROUND** Quantifying health status is becoming increasingly important for CHF. The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a new, self-administered, 23-item questionnaire that quantifies physical limitations, symptoms, self-efficacy, social interference and quality of life.
- METHODS** To establish the performance characteristics of the KCCQ, two distinct patient cohorts were recruited: 70 stable and 59 decompensated CHF patients with ejection fractions of <40. Upon entry into the study, patients were administered the KCCQ, the Minnesota Living with Heart Failure Questionnaire and the Short Form-36 (SF-36). Questionnaires were repeated three months later.
- RESULTS** Convergent validity of each KCCQ domain was documented by comparison with available criterion standards ($r = 0.46$ to 0.74 ; $p < 0.001$ for all). Among those with stable CHF who remained stable by predefined criteria ($n = 39$), minimal changes in KCCQ domains were detected over three months of observation (mean change = 0.8 to 4.0 points, $p = \text{NS}$ for all). In contrast, large changes in score were observed among patients whose decompensated CHF improved three months later ($n = 39$; mean change = 15.4 to 40.4 points, $p < 0.01$ for all). The sensitivity of the KCCQ was substantially greater than that of the Minnesota Living with Heart Failure and the SF-36 questionnaires.
- CONCLUSIONS** The KCCQ is a valid, reliable and responsive health status measure for patients with CHF and may serve as a clinically meaningful outcome in cardiovascular research, patient management and quality assessment. (J Am Coll Cardiol 2000;35:1245-55) © 2000 by the American College of Cardiology
-

Congestive heart failure (CHF) is a common clinical syndrome with an enormous impact on the prognosis and lifestyle of patients. Furthermore, it is a chronic condition without a "cure." The two primary goals of its management are preventing further disease progression (mortality, hospitalizations and deterioration of left ventricular function) and alleviating patient suffering. Quantifying this latter treatment goal requires the use of a health-related quality of life (HRQoL) instrument. In 1994, the Agency for Health Care Policy and Research published clinical practice guidelines recommending that providers routinely evaluate patients' HRQoL and use their assessment to modify and guide patient care (1). These guidelines also called for

further research to develop better mechanisms of quantifying HRQoL.

Health-related quality of life instruments may be either generic measures of health status or disease-specific. Disease-specific measures quantify more clinically relevant domains than generic health status measures and are often more sensitive to clinical change (2,3). Currently, there is no standardized, comprehensive, universally accepted disease-specific HRQoL instrument for CHF. Traditionally, a physician would use the New York Heart Association (NYHA) classification to assess functional status (a combination of physical limitations and symptoms). Although simple, this system is subject to interobserver variability, captures only a limited range of health status and is applied from a physician's perspective instead of the patient's. Other available disease-specific HRQoL instruments for CHF, such as the Minnesota Living with Heart Failure Questionnaire (4) (LiHFe) or the Chronic Heart Failure Questionnaire (5), are limited in one or more of the following: their

From the *Section of Cardiology, Department of Medicine, University of Missouri—Kansas City; and the †Mid America Heart Institute, Saint Luke's Hospital, Kansas City, Missouri.

Manuscript received August 4, 1999, accepted December 15, 1999.

Abbreviations and Acronyms

CHF	= congestive heart failure
EF	= ejection fraction
HRQoL	= health-related quality of life
KCCQ	= Kansas City Cardiomyopathy Questionnaire
LiHFe	= Minnesota Living with Heart Failure Questionnaire
NYHA	= New York Heart Association
QoL	= quality of life
SF-36	= Short Form 36
6 MWT	= 6 minute walk test

lack of responsiveness to clinical change (6–8), the range of clinical domains that they quantify, their mode of administration or their clinical interpretability. Given the breadth and complexity of quality of life (QoL) as it may be impacted upon by CHF, there is a need for improved instruments that overcome these deficiencies (1).

The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a new, self-administered, 23-item questionnaire developed to provide a better description of HRQoL in patients with CHF. It quantifies, in a disease-specific fashion, physical limitation, symptoms (frequency, severity and recent change over time), QoL, social interference and self-efficacy. Before being used, an HRQoL instrument must be rigorously evaluated and demonstrated to be valid (measure what it intends to), reliable (reproducible over time in stable patients) and responsive to clinical change. This report documents the development of the KCCQ, its validity, its reliability and its responsiveness.

METHODS

Development and design of the KCCQ. The KCCQ (Appendix 1) was developed after a thorough review of the medical literature, an examination of currently available HRQoL instruments and focus groups with patients and CHF specialists. These efforts identified the following clinically relevant domains: physical limitations (question 1), symptoms (frequency [questions 3, 5, 7 and 9], severity [questions 4, 6 and 8] and change over time [question 2]), self-efficacy and knowledge (questions 11, 12), social interference (question 16) and QoL (questions 13–15). The design and testing of the KCCQ is principally modeled after the approach used to develop the Seattle Angina Questionnaire (9). The last item of the QoL domain (question 15) was adapted from the Mental Health Inventory of the SF-36 because it is a marker of depression (10), an important prognostic variable in cardiovascular disease (11–13).

Questions were carefully formatted for gender neutrality, clarity, interpretability and subsequent foreign language translations (14). Responses were arrayed on an adjectival (Likert) scale with clinically meaningful gradations between categories. A two-week time frame was selected to allow patients to integrate a representative period of their recent

functioning into their responses. The KCCQ is self-administered and requires, on average, 4–6 min to complete. It is scored by assigning each response an ordinal value, beginning with 1 for the response that implies the lowest level of functioning and summing items within each domain. Missing values within each domain are assigned the average of the answered items within that same domain. Scale scores are transformed to a 0 to 100 range by subtracting the lowest possible scale score, dividing by the range of the scale and multiplying by 100. To facilitate interpretability, two summary scores were developed. Combining the physical limitation and symptom domains (excluding symptom stability) forms a functional status score. A clinical summary score can be calculated by combining the functional status with the QoL and social limitation domains.

Study design. Instrument selection. To place the KCCQ in the context of alternative methods of quantifying health status, two other instruments were administered. The most commonly used generic health status measure is the RAND Short Form-36 (SF-36) (15). It was scored according to the RAND criteria with each domain being transformed to a 0–100 scale where higher scores indicated better health (16). The LiHFe was also administered throughout these investigations because it is the most commonly employed disease-specific health status measure for CHF. To facilitate comparison of the LiHFe with the other instruments, its scores were reversed from the usual scoring scheme and were transformed to a 0–100 scale where higher scores indicate more favorable states. These mathematical manipulations simplify comparisons between instruments without altering the performance characteristics of the questionnaire.

Patient selection. To confirm the KCCQ's validity, reliability and responsiveness, two distinct patient cohorts were recruited, and multiple analyses were planned and conducted (Table 1). All patients had a clinical diagnosis of CHF and a documented ejection fraction (EF) of <40%. Determination of EF could have been by nuclear imaging techniques, echocardiogram or left ventriculogram. For purposes of these studies, the most recent estimation of left ventricular function was used. Patients were excluded if they could not read English. The Institutional Review Boards of the University of Missouri—Kansas City and the Mid-America Heart Institute approved all aspects of these studies.

Reliability cohort. This cohort was assembled to demonstrate the instrument's test-retest reliability. It was therefore essential to include patients who would remain stable throughout the period of observation. Requirements for stability upon entry included that over the prior two months there were: 1) no hospitalizations, 2) no change in CHF medications, and 3) a physician's assessment of stability. Patients were enrolled from the outpatient cardiology clinics of Truman Medical Center and the Mid-America Heart

Table 1. Study Designs

Study and Patients	Reference Measure	Analysis
Reliability Study Cohort with stable CHF	Three-month change in KCCQ scores	Paired <i>t</i> test
Responsiveness Study Cohort with initially decompensated CHF	Three-month change in KCCQ scores	Paired <i>t</i> test; Guyatt's responsiveness statistic
Validation of KCCQ Domains		
Physical limitation	Distance covered in a 6 minute walk test, NYHA class, SF-36 and LiHFe physical limitation domains	Correlation
Symptom stability	Difference in scores between stable and decompensated patients	Two-sample, independent <i>t</i> test
Symptoms	NYHA class	Correlation
Self-efficacy	Difference in scores between stable and decompensated patients	Two-sample, independent <i>t</i> test
Social limitation	NYHA class, SF-36 social scale	Correlation
Quality of life	SF-36 general health scale, NYHA	Correlation
KCCQ functional status	NYHA class, end point of survival or hospitalization	Correlation, <i>t</i> test
KCCQ clinical summary	NYHA class, end point of survival or hospitalization	Correlation, <i>t</i> test

CHF = congestive heart failure; KCCQ = Kansas City Cardiomyopathy Questionnaire; LiHFe = Minnesota Living with Heart Failure Questionnaire; NYHA = New York Heart Association classification; and SF-36 = Short Form-36.

Institute, Kansas City, Missouri, between September 1997 and April 1998.

Upon entry into the study, patients were administered the KCCQ, the LiHFe and the SF-36 as well as a 6-min walk test (6 MWT), a validated measure of physical function in patients with CHF (17). Patients were telephoned three months later by a cardiologist who, blinded to initial and follow-up questionnaire responses, determined their NYHA class. After this assessment, patients were mailed the KCCQ, the LiHFe and the SF-36. Patients failing to return the questionnaires were again contacted by telephone and encouraged to participate. Questionnaires were not administered by telephone. Patients who reported a significant change in their symptoms, required hospitalization or had had their CHF medicines altered during the period of observation were excluded from the reliability analyses but retained for purposes of questionnaire validation.

Responsiveness cohort. A second cohort of patients was assembled to demonstrate the instrument's responsiveness. As opposed to the reliability cohort in whom patients were selected for clinical stability, this cohort was selected to capture patients whose clinical status was expected to change. It was hypothesized that patients admitted to the hospital with decompensated CHF would be clinically improved when surveyed three months later. Patients admitted with CHF exacerbations between September 1997 and April 1998 were recruited for the study. An inability to read English was the only exclusion criterion. On admission, patients were assigned a NYHA classification and asked to complete the KCCQ, the LiHFe and the SF-36. Three months later a cardiologist, blinded to questionnaire

responses, telephoned patients and determined their NYHA functional classification. After this assessment, patients were mailed the KCCQ, the LiHFe and the SF-36. Patients failing to return the questionnaires were again contacted by telephone and encouraged to participate, but questionnaires were not administered by telephone. Per study protocol, patients who died, underwent heart transplantation or who were readmitted to the hospital were determined not to have improved over the period of observation and were excluded from the responsiveness analyses. Their baseline information was retained for instrument validation.

STATISTICAL ANALYSES

Determining questionnaire reliability. Internal consistency was determined using Cronbach's alpha. This statistic (range 0-1) reflects the "cohesiveness" of different items within a single domain. To evaluate the test-retest reliability, or reproducibility, of the instruments, paired *t* tests were used to analyze the three-month change in scores among patients retained in the reliability cohort.

Determining questionnaire responsiveness. The responsiveness of questionnaires to clinical change was quantified using paired *t* tests among patients retained in the responsiveness cohort. A responsiveness statistic was also used to compare the relative sensitivity to clinical change of the instruments used in these studies (18). This statistic divides the observed change in patients recovering from a CHF exacerbation (the "signal" of clinical improvement) by the standard deviation of change in stable patients (the "noise" of random variation). The higher the responsiveness statis-

tic, the greater the sensitivity of the instrument in detecting clinical change. Although this statistic ideally uses the minimal clinically important difference in scores as the numerator, this value is not currently known for the KCCQ. Nevertheless, it provides a useful metric with which to compare the sensitivities of different instruments.

Determining the validity of the KCCQ. Because no “gold standard” for functional status or QoL is available, convergent validity of each domain was sought by comparison with other measures that quantify similar concepts (19). Baseline data from both the reliability and responsiveness cohorts were used to assess the validity of each KCCQ domain.

Validation of the physical limitation domain. The physical limitation domain was validated against walking distances during a 6 MWT performed at baseline among patients in the reliability cohort (17). The rationale for this comparison lies in the recognition that patients reporting more physical limitation on the KCCQ would be expected to walk a shorter distance in 6 min than patients reporting less physical limitation. Distance traveled on the 6 MWT was, therefore, correlated with the physical limitation domain of the KCCQ. To further demonstrate the convergent validity of the KCCQ's physical limitation domain, it was compared with the NYHA classification and the physical limitation domains of the SF-36 and the LiHFe, other accepted measures of physical limitation.

Validation of the symptom domain. Because the NYHA quantifies both symptoms and physical limitations, it was hypothesized that a scale accurately quantifying symptoms should have worse scores for higher NYHA classes. An analysis of variance with a test for linear trend was performed to compare baseline KCCQ symptom scores with NYHA classification.

Validation of symptom stability domain. Unlike other domains of the KCCQ that quantify, in a cross-sectional fashion, patients' health status, this domain evaluates recent changes in symptoms over time. It was hypothesized that stable patients should have scores near 50 (indicating no change in symptoms over the past month) and that patients admitted to the hospital for decompensated CHF should have scores below 50, indicating recent worsening of their symptoms. The baseline mean score of this question among stable patients seen in the reliability cohort was compared with patients from the responsiveness cohort admitted to the hospital with decompensated CHF.

Validation of the self-efficacy domain. This domain captures information that is relatively unique, and we are unaware of any other validated measure that quantitatively summarizes this information. We hypothesized that patients admitted to the hospital for a CHF exacerbation would be less knowledgeable about how to manage their disease than stable outpatients. A two-sided, independent *t* test was used to compare the baseline scores of patients in the reliability and responsiveness cohorts.

Validation of the QoL domain. The QoL domain seeks to quantify how patients' disease impacts their overall QoL. This scale was correlated with the general health perception scale of the SF-36. As a post-hoc analysis, a comparison of these two domains with the NYHA classification was done to determine which was more accurately reflecting the QoL in patients with CHF. This comparison is predicated on the assumption that the NYHA classification also correlates with QoL.

Validation of the social limitation domain. This domain was correlated with the social functioning domain of the SF-36, a generic health status measure quantifying a similar concept, and the NYHA classification.

Validation of the KCCQ functional status score. We hypothesized that the functional status scores would be substantially different among patients in different NYHA classifications. The rationale for this expectation lies in the recognition that this score represents a synthesis of the physical and symptom limitations due to CHF, the very same considerations taken into account by physicians as they assign NYHA classifications. An analysis of variance with a test for linear trend was used to compare baseline functional status scores with NYHA classifications. In addition, patients who were initially enrolled in these investigations and subsequently died or were hospitalized would be expected to have worse health status at the time of their initial assessment than those with event-free survival. An independent *t* test of baseline functional status scores was performed between those who died or were hospitalized and those with event-free survival.

Validation of the KCCQ clinical summary score. In addition to the components included in the functional status score, this score also includes the social limitation and QoL domains of the KCCQ. This score was validated in the same fashion as the KCCQ functional status score by comparisons with NYHA class and by a *t* test of baseline scores among those with and without an event-free survival during follow-up procedure.

RESULTS

Summary of study cohorts. *Reliability cohort.* Seventy patients were identified as having been stable and were recruited into the reliability cohort. Over the next three months, 31 patients were removed from the cohort for the following reasons: 17 patients required hospitalization or had had their medications changed, 9 underwent heart transplantation or died and 5 did not return their follow-up questionnaires. Patients without follow-up questionnaires for analysis were not significantly different from those who did participate in follow-up with respect to their baseline characteristics (sociodemographic, clinical or health status). The mean age of the 39 stable patients available for the reliability analyses was 64 years; 69% were men, and the mean NYHA was 2.0 ± 0.59 .

Table 2. Internal Consistency of the KCCQ

KCHFQ Domain	Cronbach's Alpha
Physical limitation	0.90
Symptoms	0.88
Quality of life	0.78
Social limitation	0.86
Self-efficacy	0.62
KCCQ functional status	0.93
KCCQ clinical summary	0.95

Responsiveness cohort. Fifty-nine patients were initially recruited into the responsiveness cohort. Twenty patients were removed from the cohort for the following reasons: nine patients died or required heart transplantation, three were readmitted with CHF, three were lost to follow-up, and five failed to return the questionnaires. Patients without follow-up questionnaires for analysis were not significantly different in terms of any baseline characteristics (sociodemographic, clinical or health status) except that those without follow-up were younger than those with follow-up (57.2 vs. 67.4, $p = 0.036$). Thirty-nine patients returned the final three-month questionnaires: 62% were men; the mean age was 68 years, and the mean baseline NYHA classification was 3.3 ± 0.46 .

Validation analyses. For the validation analyses comparing the KCCQ with reference standards, all 129 patients

with available baseline data were included. The combined cohort had a mean age of 64.3 years, a mean EF of 23.5% and 70% were men. All NYHA classes were represented (I = 7; II = 41; III = 61; IV = 20).

Reliability of the KCCQ. Table 2 summarizes the internal consistency of each KCCQ domain. Only the self-efficacy domain failed to attain a high internal consistency. The lower Cronbach's alpha for this domain reflects that it is composed of only two questions acquiring slightly different pieces of information.

Table 3 summarizes the mean scores at baseline and follow-up (mean duration of follow-up = 3.3 months) among the 39 patients retained in the reliability cohort. The mean changes of 0.8 to 4.0 points on a 100-point scale over three months of observation were small and not statistically significant. Similar reproducibility was noted for both the LiHFe and the NYHA. Only the SF-36's social limitation scale changed significantly.

Responsiveness of the KCCQ. In contrast to the stability of KCCQ scores observed among members of the reliability cohort, dramatic changes were seen in the responsiveness cohort. The mean baseline, three-month and change in scores are shown in Table 4. All scales of the KCCQ were substantially more sensitive to the important clinical change experienced by this cohort, namely recovering from a severe exacerbation of CHF, than either the LiHFe or the SF-36. Whereas the KCCQ scores improved by 15.4 to 40.4

Table 3. Mean Three-Month Change in Health Status Scores Among Patients (n = 39) in the Reliability Cohort

Questionnaire and Scale	Baseline Mean Value	Three-Month Mean Value	Mean Difference	p Value
KCCQ				
Physical limitation	64.4	65.2	0.8	0.72
Symptoms	76.6	73.1	-3.5	0.06
Symptom stability	54.3	55.5	1.2	0.61
Social limitation	59.2	57.4	-1.8	0.68
Self-efficacy	83.3	85.5	2.2	0.46
Quality of life	64.5	60.5	-4.0	0.29
KCCQ functional status	70.5	69.1	-1.4	0.42
KCCQ clinical summary	66.2	64.1	-2.1	0.36
LiHFe				
Physical function	52.9	49.6	-3.3	0.35
Emotional scale	69.7	68.9	-0.8	0.78
Total LiHFe score	62.1	59.0	-3.1	0.20
SF-36				
Physical function	49.7	46.5	-3.2	0.19
Bodily pain	76.8	72.1	-4.7	0.29
General health	47.8	44.7	-3.1	0.20
Mental health	71.6	69.3	-2.3	0.28
Vitality	46.9	43.0	-3.9	0.20
Role—emotional	60.5	55.3	-5.2	0.51
Social function	76.9	65.4	-11.5	0.008
Role—physical	52.9	49.6	-3.3	0.11
NYHA	2.08	2.05	-0.03	0.57

Table 4. Mean Three-Month Change in Questionnaire Scores Among Patients Initially Admitted to the Hospital With CHF (n = 39)

Questionnaire and Scale	Baseline Mean Value	Three-Month Mean Value	Mean Difference	p Value	Responsiveness Statistic
KCCQ					
Physical limitation	34.7	55.7	21.0	<0.001	1.48
Symptoms	31.3	67.4	36.1	<0.001	3.19
Symptom stability	24.1	64.5	40.4	<0.001	2.62
Social limitation	31.1	47.9	16.8	0.004	0.62
Self-efficacy	67.6	83.0	15.4	<0.001	0.83
Quality of life	30.5	50.6	20.1	<0.001	0.86
KCCQ functional status	33.0	61.5	28.5	<0.001	2.77
KCCQ clinical summary	31.8	56.1	24.3	<0.001	1.74
LiHFe					
Physical function	32.5	43.8	11.3	0.02	0.52
Emotional scale	41.6	52.1	10.5	0.06	0.64
Total LiHFe score	38.1	49.0	10.9	0.01	0.73
SF-36					
Physical function	26.8	35.5	8.7	0.02	0.59
Bodily pain	60.4	58.6	-1.8	0.73	0.07
General health	37.2	37.4	0.2	0.98	0.01
Mental health	62.9	67.4	4.5	0.18	0.35
Vitality	27.3	38.2	10.9	0.02	0.60
Role—emotional	34.3	38.4	4.1	0.53	0.08
Social function	49.7	55.4	5.7	0.38	0.22
Role—physical	15.5	19.1	3.6	0.53	0.08
NYHA	3.2	2.4	-0.8	<0.001	2.86

points, the largest change in either the LiHFe or the SF-36 was 11.3 points. This increased sensitivity to clinical change, over the random variation seen in stable patients, is summarized by the responsiveness statistics. For example, the KCCQ physical limitation scale's responsiveness statistic of 1.48 is almost three times larger than the corresponding domains of the LiHFe (responsiveness statistic = 0.52) and SF-36 (responsiveness statistic = 0.59). The KCCQ Symptom scale has the largest "signal-to-noise" ratio (responsiveness statistic = 3.19) of any instrument used in these studies. The increased sensitivity of the KCCQ domains to clinical change, as compared with their baseline variability in stable patients, should greatly increase the KCCQ's ability to detect important clinical changes in future clinical trials.

Validation of the KCCQ. Physical limitation domain. Table 5 outlines the results supporting the validity of the KCCQ physical limitation domain. The KCCQ physical limitation score had a high Spearman's correlation coefficient with NYHA classification (-0.65, p < 0.001) and other measures of physical limitation. Among the 49 members of the reliability cohort who underwent the 6 MWT (mean age 62 years; 69% male; mean EF = 25%; NYHA I = 5, NYHA II = 28 and NYHA III = 16), the KCCQ correlated significantly with distance walked (0.48, p < 0.001).

Symptom stability score. On this scale, lower scores indicate worsening CHF symptoms, and higher scores represent symptom improvement over the preceding two weeks. A score of 50 indicates no change in symptoms. The mean KCCQ symptom stability scores were significantly lower

Table 5. Correlation of KCCQ Physical Limitation Scale

Measure	6-min Walk Test	NYHA	SF-36 Physical Limitation	LiHFe Physical
KCCQ physical limitation	0.48**	-0.65**	0.84**	0.65**
LiHFe physical	0.41*	-0.58**	0.72**	—

*p < 0.01; **p < 0.001.

KCCQ = Kansas City Cardiomyopathy Questionnaire; LiHFe = Minnesota Living with Heart Failure; NYHA = New York Heart Association; SF-36 = Short Form-36.

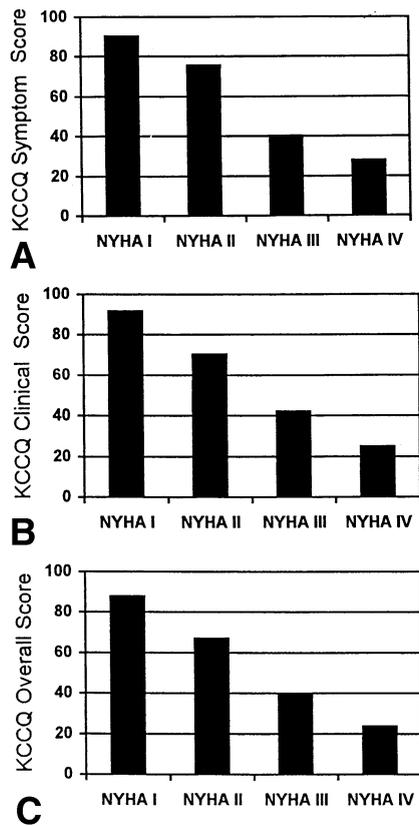


Figure 1. A. Mean KCCQ_{symptom} score by NYHA classification B. Mean KCCQ_{Functional Status} Score by NYHA classification C. Mean KCCQ_{Clinical Summary} Score by NYHA classification

among patients admitted to the hospital (n = 59) than among those who were stable (n = 70) (25.8 vs. 53.8, p < 0.0001).

Symptom scale. This scale summarizes the results of the symptom frequency and symptom severity domains of the KCCQ. Figure 1A summarizes the mean scores among patients in different NYHA classes. An analysis of variance revealed statistical differences (F = 51.3, p < 0.001) that were largely explained by a linear trend (R² = 0.51; F = 142.2, p < 0.001). Similar relationships were observed between the individual components of the symptom score (frequency and severity) and NYHA classification when analyzed individually.

QoL domain. A significant correlation (r = 0.45, p < 0.001) was seen between the KCCQ_{QoL} domain and the general health perception scale of the SF-36. Since the correlation with the SF-36 scales was only modest, it was unclear which of the scales was more accurately capturing QoL among patients with CHF. A much higher correlation with NYHA classification was seen with the KCCQ (r = -0.64, p < 0.001) than with the SF-36 general health scale (r = -0.46, p = 0.02). The KCCQ_{QoL} domain also correlated significantly with the emotional domain of the LiHFe (r = 0.62, p < 0.001).

Social limitation domain. The KCCQ_{social limitation} domain was significantly correlated with the SF-36 social limitation scale (r = 0.62, p < 0.001) and with NYHA class (r = -0.57, p < 0.001).

Self-efficacy domain. No adequate criterion standard is available for this domain, and acceptance of this domain could be made on its “face” validity, reliability and responsiveness. To further justify the validity of these items, a comparison was made between the self-efficacy of stable outpatients with CHF and those who are admitted to the hospital with decompensated symptoms. Tables 3 and 4 demonstrate that scores were significantly lower among patients admitted to the hospital than among stable outpatients (67.6 vs. 83.5, p < 0.0001). Among patients in the responsiveness cohort who improved, self-efficacy scores also improved significantly at follow-up (67.6 vs. 83.0, p < 0.001). This may be due to patients’ better understanding of how to manage their disease as a result of the educational processes that occur during hospital admission at the two recruiting centers.

KCCQ_{functional status} score. As shown in Figure 1B, mean scores on this scale were different among patients in different NYHA classifications (F = 52.3, p < 0.001). Not only was this scale closely correlated with NYHA classification (test for linear trend F = 153 and R² = 0.55; p < 0.001), but baseline scores were also associated with subsequent death or hospitalization during follow-up procedures. Over the course of the study, 11 patients died and 13 patients required hospitalization. The baseline KCCQ_{functional status} score was significantly lower among patients dying or requiring rehospitalization during the follow-up period than among those with event-free survival (35.1 vs. 55.3, p < 0.001).

KCCQ_{clinical summary} score. This scale includes the functional status score as well as the KCCQ_{QoL} and social limitation scores. Like the functional status score, the functional status score correlated with NYHA class (Fig. 1C, overall analysis of variance F statistic = 41.9, p < 0.001; test for linear trend F = 156.8, p < 0.001; R² = 0.55). There also was a significantly lower baseline score among patients who died or were rehospitalized than those with event-free survival (34.1 vs. 52.1, p = 0.001).

DISCUSSION

In an era of rapidly expanding therapies for CHF, there is a need for robust mechanisms of quantifying the impact of new treatments on patients, their survival, their symptoms and their QoL. In addition, payers, practitioners and regulatory agencies are increasingly relying upon patient-centered outcomes to monitor and improve quality of care. To meet these needs, we have developed a new, self-administered, disease-specific health status measure for patients with CHF.

This report details the development, validity, reliability and reproducibility of the KCCQ. In support of its validity,

predictable and significant relationships with the best available criterion standards were demonstrated for each individual domain of the KCCQ. Among a cohort of stable patients with CHF, no significant changes were detected over a three-month period of observation. In contrast, large changes were observed among a separate cohort of patients recovering from an admission for decompensated heart failure.

The KCCQ offers several advantages over the LiHFe, another disease specific HRQoL instrument for CHF. The KCCQ not only captures physical limitations, but also independently quantifies symptoms (frequency, severity and stability), social limitations, patients' sense of self-efficacy and QoL. Although both the KCCQ and LiHFe appear to be valid and reliable, the KCCQ is much more sensitive to clinical change than the LiHFe. The lack of sensitivity to clinical change may, in part, explain why some clinical trials have failed to demonstrate important benefits in QoL despite improvements in other clinical markers (6-8). The increased sensitivity of the KCCQ was also apparent when compared with the SF-36. This is consistent with the growing evidence of generic health status instruments' limited responsiveness to treatment effects (3,20,21).

The Chronic Heart Failure Questionnaire developed by Gordon Guyatt (5) is a reliable and responsive measure for patients with CHF (22). However, its usefulness has been limited by the complexity of its administration.

Study limitations. Validating the KCCQ was complicated by the absence of appropriate reference standards for the various domains. Although we recognize the limitations of our comparison standards, we feel these were the best options available. A second concern is the loss to follow-up of four patients (3%) and the failure of nine patients (7%) to

return follow-up questionnaires. Although it is unknown how their responses might have altered our findings, it seems unlikely that they would significantly alter our conclusions. Furthermore, future users of our instrument are likely to have a portion of their subjects similarly fail to provide follow-up, and our results are likely to be generalizable to those of their patients that do provide follow-up responses.

Finally, we recognize that we have examined this instrument in a relatively small number of patients. Further work in other populations, including those with diastolic rather than systolic dysfunction, will better define the generalizability of our findings. In addition, we have not yet determined the minimal clinically important difference in KCCQ scores. More experience in different clinical settings should further illuminate this information.

Potential applications of the KCCQ. It is hoped that the KCCQ will serve as an end point in clinical trials and be useful in disease management or quality assessment/improvement programs. In addition to mortality, investigators often rely on physiologic variables, such as left ventricular EF or plasma norepinephrine levels as surrogate end points. However, surrogate end points may not always be as useful as expected (23). The KCCQ can directly quantify clinical benefits from patients' perspectives.

Conclusions. The benefit and importance of functional status assessment and patient outcomes data has become increasingly recognized. The KCCQ is a valid, reliable and responsive HRQoL instrument for patients with CHF and may serve as a clinically meaningful outcome measure in cardiovascular research, patient management and quality assessment.

Acknowledgments

The authors gratefully acknowledge the instrumental contributions of the following towards the completion of this project: Brent Bliven, MSc, PhD; Nancy D. Long, RN; Shamila Garg, MD; Annette Quick, MD; Steve Freeman, MD; Elaine Russell, RN; Ann Stewart, RN; Kathy St. Claire, RN; Marcia McCoy, RN, MSN; and Elly Lensch.

Reprint requests and correspondence: Dr. John A. Spertus, Mid America Heart Institute/University of Missouri—Kansas City, 4401 Wornall Road, Kansas City, Missouri 64111. E-mail: jspertus@cctr.umkc.edu.

REFERENCES

1. Konstam M, Dracup K, Baker D, et al. Heart Failure: Evaluation and Care of Patients with Left Ventricular Systolic Dysfunction. AHCPR Publication No. 94-0612, 1994.
2. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
3. Spertus JA, Winder JA, Dewhurst TA, Deyo RA, Fihn SD. Monitoring the quality of life in patients with coronary artery disease. *Am J Cardiol* 1994;74:1240-4.
4. Rector T, Kubo S, Cohn J. Patient's 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.
5. Guyatt GH, Nogradi S, Halcrow S, Singer J, Sullivan MJ, Fallen EL. Development and testing of a new measure of health status for clinical trials in heart failure. *J Gen Intern Med* 1989;4:101-7.
6. Colucci W, Packer M, Bristow M, et al. Carvedilol inhibits clinical progression in patients with mild symptoms of heart failure. *Circulation* 1996;94:2800-6.
7. Packer M, Colucci W, Sackner-Bernstein J, et al. Double-blind, placebo controlled, study of the effects of carvedilol in patients with moderate to severe heart failure. *Circulation* 1996;94:2793-9.
8. Bristow M, Gilbert E, Abraham W, et al. Carvedilol produces dose-related improvements in left ventricular function and survival in subjects with chronic heart failure. *Circulation* 1996;94:2807-16.
9. Spertus JA, Winder JA, Dewhurst TA, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *J Am Coll Cardiol* 1995;25:333-41.
10. Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein M. Performance of a five-item mental health screening test. *Med Care* 1991;29:169-76.
11. Frasure Smith N, Lesperance F, Talajic M. Depression following myocardial infarction. Impact on six-month survival. *JAMA* 1993;270:1819-25.
12. Frasure Smith N, Lesperance F, Talajic M. Depression and 18-month prognosis after myocardial infarction. *Circulation* 1995;91:999-1005.
13. Vaccarino V, Kasl SV, Krumholz HM. Effect of depression on mortality and functional status among patients with decompensated heart failure. *Circulation* 1998;98:I-753.
14. Spertus J, Green C, Conway K, Uzun V. Creating a new disease-specific health status measure for congestive heart failure: a prototype for instrument development in the current era. *Quality of Life Newsletter* 1998;20:9-10.
15. McHorney C, Ware J, Raczek A. The MOS-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247-63.
16. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd ed. New York: Oxford University Press, 1996.
17. Guyatt G, Sullivan M, Thompson P, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Mes Assoc J* 1985;132:919-23.
18. Deyo R, Dieher P, Patrick D. Reproducibility and responsiveness of health status measures. *Controlled Clinical Trials* 1991;12:142S-58S.
19. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press, 1989.
20. Weinberger M, Oddone EZ, Henderson WG. Does increased access to primary care reduce hospital readmissions? Veterans Affairs Cooperative Study Group on Primary Care and Hospital Readmission. *New Engl J Med* 1996;334:1441-7.
21. Ekeberg O, Klemsdal TO, Kjeldsen SE. Quality of life on enalapril after acute myocardial infarction. *Eur Heart J* 1994;15:1135-9.
22. Guyatt GH, Sullivan MJ, Fallen EL, et al. A controlled trial of digoxin in congestive heart failure. *Am J Cardiol* 1988;61:371-5.
23. Fleming TR, DeMets DL. Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 1996;125:605-13.