MINI-FOCUS ISSUE: PATIENT-REPORTED OUTCOMES

STATE-OF-THE-ART PAPER

Utility of Patient-Reported Outcome Instruments in Heart Failure

Anita A. Kelkar, MD, MPH,[a] John Spertus, MD,[b] Peter Pang, MD,[c] Renee F. Pierson, MBA,[d] Robert J. Cody, MD, MBA,[d] Ileana L. Pina, MD,[e] Adrian Hernandez, MD,[f] Javed Butler, MD, MPH[g]

ABSTRACT

Patient-reported outcomes (PRO) are defined as reports coming directly from patients about how they feel or function in relation to a health condition and its therapy. Although there are numerous compelling reasons why PRO could be an important help in clinical care, they have not evolved into clinical tools integrated into care. The purpose of this review is to assess existing PRO instruments for heart failure with respect to their psychometric properties and potential for use in clinical care. We performed a systematic search of articles published between July 2008 and January 2015 within the MEDLINE, PROMIS, PROQOLID, and Cochrane Library databases. Included instruments had to be developed and tested for heart failure and have had their development processes and psychometric properties described. A total of 31 instruments were identified, 9 of which met all inclusion criteria. After evaluating each remaining instrument in terms of psychometric and clinical criteria and symptom coverage, only 2 instruments—Minnesota Living with Heart Failure and Kansas City Cardiomyopathy questionnaire—met all evaluation criteria. Although clinically useful PRO instruments exist, increasing education to providers on the value and interpretability of PRO instruments, as well as a more streamlined approach to their implementation in the clinical setting is necessary. A clinical trial comparing the routine use of disease-specific PRO with clinical care could further support their incorporation into practice. (J Am Coll Cardiol HF 2016;4:165–75) © 2016 by the American College of Cardiology Foundation.

Highlighted by the Institute of Medicine's *Crossing the Quality Chasm* report (1) and re-emphasized by President Obama's 2015 State of the Union Address, there is an increasing need to improve the patient-centeredness of health care. Heart failure (HF) is an example of an entity in which patient adherence to optimal medical therapy as well as patients understanding their symptoms can help guide management. Furthermore, with the increasing use of invasive but life-sustaining technology, such as left ventricular assist devices or transaortic valve replacement, the physician's understanding of patient's goals and values are essential for HF management and decision making (2). A critical

step in this effort is to be able to validly, reproducibly, and sensitively measure patients' experiences with their illness. Patient-reported outcomes (PRO) have been proposed as such a method. In this review, we discuss the history of PRO use in HF, provide an assessment of existing PRO instruments with respect to their psychometric properties, and discuss the clinical landscape of PRO use in HF with suggestions for which PRO could be potentially used in clinical care as well as clinical trials.

We focus on disease-specific PRO measures because such tools are more sensitive to clinical change than generic health status measures are (3), which is particularly important when considering the use of these instruments in clinical care. Although there have been at least 2 previous reviews (4,5) of disease-specific PRO in HF, this review focuses attention on the use of these measures for clinical care, as opposed to use in clinical trials. This application of PRO is particularly relevant given emerging interests in having PRO transition into performance measures of quality (6). Whereas it is potentially important to use assessments of patients' health status as markers of health care quality, reporting these assessments is a substantial burden, unless they are a by-product of clinical use. A recent editorial (7) highlights the importance of incorporating PRO into practice. If clinicians had a deeper understanding of the potential benefits of using PRO in clinical care, it might support their acceptance and use. This review was developed to meet this need.

THE HISTORY OF PATIENT-REPORTED OUTCOMES IN CLINICAL PRACTICE

PRO are defined as reports coming directly from patients about how they feel or function in relation to a health condition and its therapy, without interpretation by health care providers or anyone else (8). PRO can provide information on a range of patients' health status outcomes including the following: symptoms; functional limitations; impacts on daily activities; social, emotional, and psychological outcomes; and overall well-being. General health status measures have been the more commonly used PRO as they measure health status across a wide range of diseases and patient types. Comparatively, HF-specific PRO are a newer entity that has seen increasing development and testing recently. Preceding HF PRO, clinician-assessed measures such as New York Heart Association (NYHA) functional class had been

primarily used. However, PRO are more reproducible than clinician-assessed symptoms or even more "objective" tests such as assessments of ejection fraction or valve gradients (9,10). Whereas it is common in practice to see different physicians assign different NYHA functional classes to the same patient, PRO can overcome this limitation by asking patients relevant questions about their health in the same way each and every time. This renders the PRO a "standardized history" and the completeness of the PRO in capturing the relevant manifestations of HF can guide its potential usefulness in clinical care. It can also overcome differences in physician- and patient-reported symptom burden, given that physicians may underestimate the impact of symptoms on patients' physical and social function (11). These are especially important considerations among patients with HF who almost universally have variable limitations in functional capacity and health status.

HF-specific PRO have been increasingly used in multiple settings including but not limited to post-hospitalization follow-up visits, regular outpatient appointments, and even during a hospitalization. Developed in the mid-1980s, the Minnesota Living with Heart Failure Questionnaire (MLHFQ) is considered to be among the first HF PRO and is among the most widely used after being validated for use as an outcome in clinical trials (12). The Kansas City Cardiomyopathy Questionnaire (KCCQ) was developed almost a decade later and is now the most widely used HF PRO and has recently been shortened to a 12-item version to support its use in clinical care (13). Despite the existence of these PRO for more than 2 decades, they are still not routinely integrated in clinical care. There are reasons to believe that PRO would be an important metric for high-quality care, including their reproducibility over time and across providers, their sensitivity to changes (improvements or deteriorations), and their prognostic significance. However, there are also challenges in the routine use of PRO, including the feasibility of implementation into clinical workflow, integration into electronic medical records, interpretability of the results, and reimbursement to support the infrastructure to collect, score, and analyze them. Ideal health status measures for HF patients should also be appropriate for the numerous etiologies of HF, and the results of the measures should be clinically actionable and potentially amenable to change with therapy. Interestingly, as the field of PRO assessment has evolved, PRO have been increasingly included as endpoints in clinical trials, but the opportunity to include them in routine clinical practice is virtually untapped (14,15).
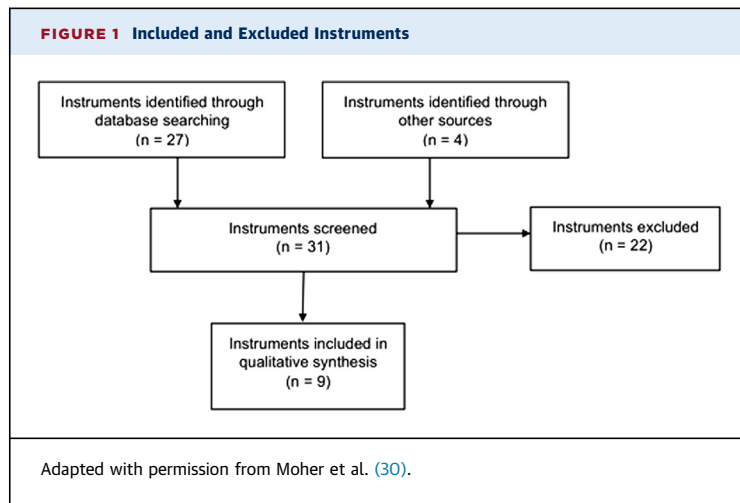
PATIENT-REPORTED OUTCOMES IN CLINICAL TRIALS

PRO have been increasingly used in clinical trials. Because PRO are often questionnaires completed by patients, they have been previously considered as "soft" endpoints. However, it should be noted that the "hardness" of data has been defined as its reproducibility (16), and PRO have shown to be more reproducible than other clinical trial measures such as assessments of ejection fraction or valve gradients (9,10). Given this, in 2009 the U.S. Food and Drug Administration created a document, *Guidance for Industry, Patient Reported Outcome Measures: Use in* *Medical Product Development to Support Labeling Claims* (17), that gives guidance on how to use PRO in the clinical trial framework as well as a potential outcomes to support labeling claims for medications. This acceptance by the U.S. Food and Drug Administration has helped propel the use of PRO in the clinical trials (18). However, consistent PRO use within clinical trials continues to lag. A systematic review published in 2010 (19) assessed the role of PRO within contemporary cardiovascular clinical trials. Of the 413 trials evaluated, approximately 65 trials (16%; SE 2%) used at least 1 PRO; many trials in which PRO may have assisted or been crucial for decision making did not report such outcomes (122 of 174 trials, 70%) (19).

TABLE 1   Measurement Properties for PRO

	Property Type	Property Description	Property Presentation	Acceptability
Reliability	Test/retest or intra-/interviewer reliability	Score stability over time when no change is expected in the concept of interest	ICC range 0–1.0	ICC reliability score >0.5
	Internal consistency	Extent to which items comprising a scale measure the same concept	Cronbach alpha Item-total correlations range 0–1.0	Cronbach alphas for summary scores ≥0.70 for group comparisons
Validity	Content validity	Instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population, and use	Not applicable	Patients of interest (HF) involved in the development stage and item generation
	Construct validity	Evidence that relationships among items, domains, and concepts conform to a priori hypotheses concerning logical relationships that should exist with measures of related concepts or scores produced in similar or diverse patient groups	Strength of correlation testing a priori hypothesis, range 0.0–1.0	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation. The strength of the correlation depends upon the similarity of the concept quantified by the PRO and the comparison standard
Responsiveness (ability to detect change)		Evidence that a PRO instrument can identify differences in scores over time in individuals or groups who have changed with respect to the measurement concept	Effect size statistic	Statistically significant differences between known groups and/or a difference of expected magnitude
Diversity in performance		Evidence that the PRO instrument is applicable to all HF populations regarding of HF etiology (e.g., ischemic cardiomyopathy, diastolic HF)	Evidence of published use in distinct HF populations	Psychometric properties of responsiveness, validity, and reliability maintained across all HF populations
Feasibility		The PRO should be clinically practical to administer in routine care. It should be self-administered and not overtly time-consuming for completion	Time to completion, need for interviewer to administer, number of questions to complete	Clinically practical time for completion (≤10 min)
Interpretability		PRO should be easy to interpret for the interpreting provider, nurse practitioner, or other relevant health care professional. PRO scores should be easy to translate to differentiate among small, moderate, and large clinical changes	Insights available to interpret scores, and changes in scores, into a clinically meaningful framework	Able to easily interpret immediately in a clinical setting without need for other resources or assistance. Scores should be easily interpreted to detect changes in clinical status and prognosis
Prognostic value		The conclusions obtained from the PRO hold prognostic value and potentially change a clinician's treatment approach	Evidence of published prognostic value in the literature	Presence of prognostic value of conclusions obtained from the PRO

HF = heart failure; ICC = interclass correlation coefficient; PRO = patient-reported outcome(s).

168   Kelkar et al.
Utility of Patient-Reported Outcome Instruments in HF

JACC: HEART FAILURE VOL. 4, NO. 3, 2016
MARCH 2016:165–75

FIGURE 1   Included and Excluded Instruments

Adapted with permission from Moher et al. (30).

It is hopeful that increased use of PRO within trials and subsequently as labeling claims for medications may also incite increased use within the general clinical arena.

ASSESSMENT OF CURRENT HEART FAILURE PATIENT-REPORTED OUTCOMES INSTRUMENTS

METHODS. Search strategy. A previous review by the PRO Measurement Group (4) identified PRO that were assessed for their development and measurement properties and applicable to HF patients. In order to identify new PRO instruments developed since this 2009 review, we performed a systematic search of articles published between July 1, 2008 and January 31, 2015. We searched the MEDLINE,

TABLE 2   Instrument Content Coverage

	CHFQ	KCCQ	MLHFQ	CHAT	LVD-36	SDHFQ	HFFSI	QLQ-SHF	MSAS-HF
Potential symptom endpoints									
Shortness of breath (dyspnea) when you exert yourself or when you lie down	x	x	x		x	x	x	x	x
Fatigue and weakness	x	x	x			x	x	x	x
Swelling (edema) in your legs, ankles, and feet		x	x		x	x			
Rapid or irregular heartbeat									x
Reduced ability to exercise		x				x			
Persistent cough or wheezing with white or pink blood-tinged phlegm									x
Swelling of your abdomen (ascites)*									
Sudden weight gain from fluid retention*									
Lack of appetite and nausea					x				x
Difficulty concentrating or decreased alertness			x	x					
Chest pain						x	x	x	
Palpitation							x		
Dizzy/lightheaded						x			x
Sleeping changes/difficult/limited	x	x	x	x		x			x
Eating/drinking limited			x	x					
Frustrated or impatient	x				x				x
Feel inadequate, worthless, burden to others	x		x		x				
Low in energy/tired/fatigue	x	x	x	x	x				x
Feel upset, worried, depressed	x	x	x					x	x
Feel relaxed and free of tension	x							x	
Feel happy, satisfied, or pleased	x	x						x	
Feel restless, tense, or uptight	x							x	
Limited physically		x			x				
Limited ability to do what you wanted		x			x				
Limited enjoyment of life		x			x				
Limited hobbies, recreational activities		x	x		x				
Limited working or doing household chores		x	x		x	x	x		
Limited visiting family or friends out of home		x	x	x					
Limited intimate/sexual relationship		x	x	x		x			x
Total symptom endpoints met	10	14	13	6	11	9	5	7	11

*Symptom endpoint not covered by any instrument.

CHAT = Chronic Heart Failure Assessment Tool; CHFQ = Chronic Heart Failure Questionnaire; HFFSI = Heart Failure Functional Status Inventory; KCCQ = Kansas City Cardiomyopathy Questionnaire; LVD-36 = Left Ventricular Dysfunction Questionnaire; MLHFQ = Minnesota Living with Heart Failure Questionnaire; MSAS-HF = Memorial Symptom Assessment Scale—Heart Failure; QLQ-SHF = Quality of Life Questionnaire for Severe Heart Failure; SDHFQ = San Diego Heart Failure Questionnaire.

PROMIS, PROQOLID, and Cochrane Library databases (Online Appendix 1). All previous PRO instruments mentioned in the 2009 review were also searched using the same databases and time frame.

Inclusion and exclusion criteria. Studies could be cross-sectional or longitudinal, but they had to study HF patients and be reported in English. The studies had to measure HF symptoms and describe the development process and psychometric properties (Table 1). Only instruments developed and tested for HF were included, while generic instruments as they are less sensitive to HF changes than HF-specific instruments (3), were excluded. The instruments also had to measure patient-reported, as opposed to only clinician-reported, outcomes. Due to the unique health-related quality of life (QoL) aspects of patients with implanted mechanical ventricular assist devices, we excluded literature involving this patient population. We also excluded letters, editorials, commentary, practice guidelines, and primary discussion papers (Figure 1).

Evaluation process. Each identified instrument was evaluated according to established psychometric criteria (20,21). "Content validity" ensures that the PRO measures the concept of interest for the intended population (i.e., HF patients were involved in the development and item generation). "Construct validity" ensures that there is evidence that relationships should exist with measures of HF-related concepts and scores. "Reliability" ensures that the score obtained remains stable when patients' clinical status is unchanged, and "Responsiveness" ensures that the PRO changes when patients' conditions change and is sensitive to the magnitude of those clinical changes. These criteria were selected based on the Food and Drug Administration guidance in regard to PRO measure development (17) (Table 1). "Performance diversity" assessed whether the instrument is applicable across populations regardless of etiology. "Feasibility" assesses the practicality of use in the clinical setting, with attention to the response burden so that the PRO is not to be overly time-consuming to implement. "Interpretability" establishes magnitudes of change that are clinically important. The PRO should be straightforward to interpret by the provider who should be able to understand what amount of changes in scores correlate to large, moderate, or small changes in clinical status. PRO should also have "Prognostic Value" to enable risk stratification and tailoring of treatment to risk. The relevant articles for each instrument were evaluated for whether there was evidence to fulfill these criteria. Lastly, the questions and concepts of each PRO instrument were cross-matched with a list of

TABLE 3 Instruments Included and Excluded	
Included Instruments	
CHFQ	HFFSI
KCCQ	QLQ-SHF
MLHFQ	LVD-36
CHAT	MSAS-HF
SDHFQ	
Excluded Instruments	
Zung Self-Rating Depression Scale	Patient-Generated Index
Profile of Mood State	London Chest Activity of
Generic QoL Questionnaire	Daily Living Scale
Cumulative Illness Rating Scale	Health Utilities Index Mark-3
Single Item Health Status	Patient Health
Multidimensional Fatigue Inventory	Questionnaire-9
Social Provisions Scale	Brief Symptom Inventory
EuroQol Dimensions Index	MacNew
Medical Outcomes Study 36-Item	Cardiac Depression Scale
Short-Form Health Survey	Duke Activity Status Index
12-Item Short-Form Health Survey	QoL Index—Cardiac Version
Sickness Impact Profile	Dyspnea Fatigue Index
	Health-Related QoL
	Questionnaire

EuroQoL = European Quality of Life; QoL = quality of life; other abbreviations as in Table 2.

HF-specific symptom endpoints (Table 2). Each PRO was evaluated for whether it addressed HF as well as other detrimental symptoms that HF patients routinely experience.

RESULTS. Instruments characteristics. Table 3 depicts 9 of the 31 instruments developed for HF that fit the inclusion criteria. Additional instruments were excluded due to lack of published psychometric properties in HF patients or involvement of other measures unrelated to HF. These instruments covered a variety of themes including physical activity, social limitations, life satisfaction, and specific symptoms. They were developed using a variety of techniques, including unstructured interviews and focus groups and examination of previous or existing instruments. Some PRO could be self-administered, whereas others were designed for interview administration. Published reports of the time required for patients to complete the instruments ranged from 4 to 20 min. The median item per PRO was 26 items (range 16 to 51). Most PRO items were scored on a Likert scale, with a few either having dichotomous or numerical scales.

Evaluation criteria. Table 4 presents a summary of the details of each instrument, as well as whether or not they fulfilled the evaluation criteria. Of the 9 instruments studied, only 2 fit all of the evaluation criteria—the KCCQ and the MLHFQ.

Psychometric criteria. The KCCQ and MLHFQ had content validity by involving HF patients during development and construct validity by showing correlations between the instrument and other measures

Kellar et al.
Utility of Patient-Reported Outcome Instruments in HF

JACC: HEART FAILURE VOL. 4, NO. 3, 2016
MARCH 2016:165–75

TABLE 4 Instrument Content and Properties

	Content Covered	Content Validity	Construct Validity	Reliability	Responsiveness	Performance Across HF Populations	Feasibility	Interpretability	Prognostic Value	Number of Studies Identified for Appraisal
CHFQ	It contains 4 subscales: daily activities; fatigue; emotional function; and mastery (31).	The initial 123 items were generated through a review of literature, consultation with experts, and unstructured interviews with patients (31).	Yes (31)	Yes (32)	Yes (33)	Used in studies in HF patients of all etiologies (34).	Trained interviewer required to administer. 16 items with 7 point Likert scale. Completion time 10–20 min (31).	Scores for questions within each dimension added to give worst and best function information for each dimension (higher numbers equal worse function). A half-point difference is considered clinically significant (31).	No prognostic information available.	18
KCCQ	Consists of 6 domains and 2 summary scores: physical limitation; symptom; symptom stability; self-efficacy; QoL; social limitation; KCCQ clinical summary; and KCCQ overall summary.	Initial item generation was based on literature review, examination of HRQOL instruments, and focus groups with HF patients and specialists (22).	Yes (22)	Yes (35)	Yes (36)	Used in studies in HF patients of all etiologies. Validation work performed in HFpEF and aortic stenosis (20,37).	Self-administered. 23 items with 6 point Likert scale. Completion time 4–6 min (22). Recently a 12-item version has been introduced, taking 2–3 min (23).	Each response is assigned a value; lower scores indicate lower level of functioning. Summary scores developed to mirror NYHA (clinical summary score) and all health status domains (overall summary score) (22). Clinically significant differences that are small, moderate, and large = 5, 10, and 20 points (3). A 5-point change is associated with a fully adjusted 10% change in the hazard of death and rehospitalization (24).	Low score found to be an independent predictor of poor prognosis in outpatients with HF (38).	250
MLHFQ	Covers 3 domains: physical; socioeconomic; and psychological impairments (12).	The items were identified from a list of sickness-related dysfunction on the SIP. Reliability and validity testing in HF patients during development (12).	Yes (39)	Yes (40)	Yes (41)	Used in studies in HF patients of all etiologies (42,43).	21 items with 6-point Likert scale. Self-administered or interview. Completion time 10 min (12).	Single summation of responses, the higher the summed score, the worse the patient's QoL (12). Clinically significant differences between NYHA I, II, and III, but it does not do well in detecting subtle changes between NYHA III and IV (25). A 5-point increase in score, interpreted as a clinically improved difference in score (44).	Independent predictor of cardiac events, death, and future hospitalizations (45).	380
CHAT	Covers 5 domains (symptom activity levels, psychosocial aspects, and emotions) and covers 7 themes of HRQOL (physical ability, emotional state, self-perception, relationships, coping with symptoms, maintaining social/lifestyles status, and cognitive aspects of CHF) (46).	The measure was based on patient input. The initial 51 items were generated using in-depth, semistructured interviews with CHF patients (46).	Yes (46)	Yes (46)	No (inadequate data).	Used in studies in HF patients of all etiologies (46).	51 items with 5-point scale. Self-administered (46). Approximate time to completion 15–20 min.	Higher summed score indicates worse QoL (46). No published data on the clinical significance of a given change in score.	No prognostic information available.	2

TABLE 4 Continued

	Content Covered	Content Validity	Construct Validity	Reliability	Responsiveness	Performance Across HF Populations	Feasibility	Interpretability	Prognostic Value	Number of Studies Identified for Appraisal
SDHFQ	Content covers dyspnea and fatigue at rest and in relation to physical activities. Also includes questions on chest pain, alcohol consumption, and weight (47).	The original development article did not provide detail regarding the development or testing of the SDHFQ (47).	Yes (48)	No (inadequate data)	No (48)	Used in studies in HF patients of all etiologies (48).	32 items on a 2- to 6-point numerical scale (47). Self-administered. Approximate time to completion 10 min.	Functional capacity scored from certain questions. Each letter has a specific numerical value that depends on the question. No published data on the clinical significance of a given change in score.	No prognostic information available	3
HFFSI	Content covers overall functional status level and the frequency and type of symptoms that limit physical function (49).	The measure was adapted from a standardized self-report questionnaire that was modified for HF patients (49).	Yes (49)	Yes (50)	No (51)	Used in studies in HF patients of all etiologies (50).	25 items on a 3-point scale. Self-administered (49). Approximate time to completion 5–7 min.	Each item/activity assigned a value per the MET of the task. HFFSI score correlates to the average of the 3 highest MET levels (49). No published data on the clinical significance of a given change in score.	No prognostic information available	12
QLQ-SHF	Covers the following 5 domains: somatic; emotional; cognitive aspect; life satisfaction; and physical limitation (52).	The initial 90 items were derived from the work on infarction patients, the results from other studies, and the content of other scales (52).	Yes (52)	Yes (52)	Yes (53)	Used in studies in HF patients of all etiologies (52,53).	26 items using the Likert scale. Self-administered (52). Approximate time to completion 7–10 min.	Score uses the numerical value on the Likert scale; the higher the score, the worse the QoL. No published data on the clinical significance of a given change in score.	No prognostic information available	3
LVD-36	Covers domains of physical limitations, life satisfaction, and emotional well-being (54).	Developed using semistructured interviews with HF patient and used information from reviews of published reports and existing questionnaires (54).	Yes (54)	Yes (54)	Yes (55)	Used in studies in HF patients of all etiologies (54,56).	36 items with dichotomous answers (yes or no) measuring general health status. Self-administered. Approximately 5 min to complete (54).	Responses are summed, and sum is expressed as a percentage such that 100 is the worst score. No published data on the clinical significance of a given change in score.	No prognostic information available	7
MSAS-HF	Covers 3 symptom subscales: physical; emotional; and HF-specific symptoms (54).	Modified from the original Memorial Symptom Assessment Scale (a scale initially created for QoL in the cancer population) (57). 5 HF-specific symptoms added and 5 cancer symptoms removed after being tested in HF patients (58).	Yes (58)	Yes (58)	Yes (59)	Used in studies in HF patients of all etiologies (58).	32 items on a 4- to 5-point scale. Self-administered. Approximately 5 min to complete (56).	Score summary gives prevalence. Score mean gives burden of HF (56). No published data on the clinical significance of a given change in score.	No prognostic information available	4

CHF = chronic heart failure; HFpEF = heart failure with preserved ejection fraction; HRQOL = Health-Related Quality of Life; MET = metabolic equivalent; NYHA = New York Heart Association functional class; SIP = Sickness Impact Profile; other abbreviations as in Tables 1 to 3.

Kelkar et al.
Utility of Patient-Reported Outcome Instruments in HF
171

such as the 6-min walk test and peak exercise oxygen consumption. They demonstrated responsiveness by their ability to detect changes and differences between study groups over time. Both instruments have been used in diverse populations regardless of etiology, with the psychometric properties of the KCCQ being explicitly demonstrated in patients with HF and preserved ejection fraction and severe aortic stenosis (20,21).

Clinical criteria. The KCCQ and MLHFQ were feasible as they could be self-administered in under 10 min (12,22), with the short-form KCCQ-12 requiring only 2 to 3 min (13,23). KCCQ provides both a clinical summary score, to summarize the physical activity and symptom domains captured in the NYHA, as well as an overall summary score. KCCQ is also unique in that it has explicitly assessed the magnitude of changes in scores associated with different magnitudes of clinical change (KCCQ overall summary scores: small clinically important change, 5 points; moderate change, 10 points; and large change, 20 points or greater) (3). A 5-point change in KCCQ score was also associated with a fully adjusted 10% change in death or hospitalization risk (24). MLHFQ scores have also correlated with changes in functional classes, but they were limited in detecting differences among patients with more advanced HF (i.e., between NYHA functional classes III and IV) (25). Both the MLHFQ and KCCQ have prognostic significance, which may aid in their clinical interpretation and decision making based on risk stratification.

Although all of the other instruments had construct validity and evidence of performance across HF populations, they did not add prognostic value and/or did not meet other criteria, including content validity (San Diego Heart Failure Questionnaire [SDHFQ], Heart Failure Functional Status Inventory [HFFSI], Quality of Life Questionnaire for Severe Heart Failure [QLQ-SHF]), reliability (SDHFQ), responsiveness (SDHFQ, HFFSI), feasibility (Chronic Heart Failure Questionnaire secondary to interviewer requirement), or interpretability (Chronic Heart Failure Assessment Tool, SDHFQ, HFFSI, QLQ-SHF, Left Ventricular Dysfunction Questionnaire, Memorial Symptom Assessment Scale–Heart Failure).
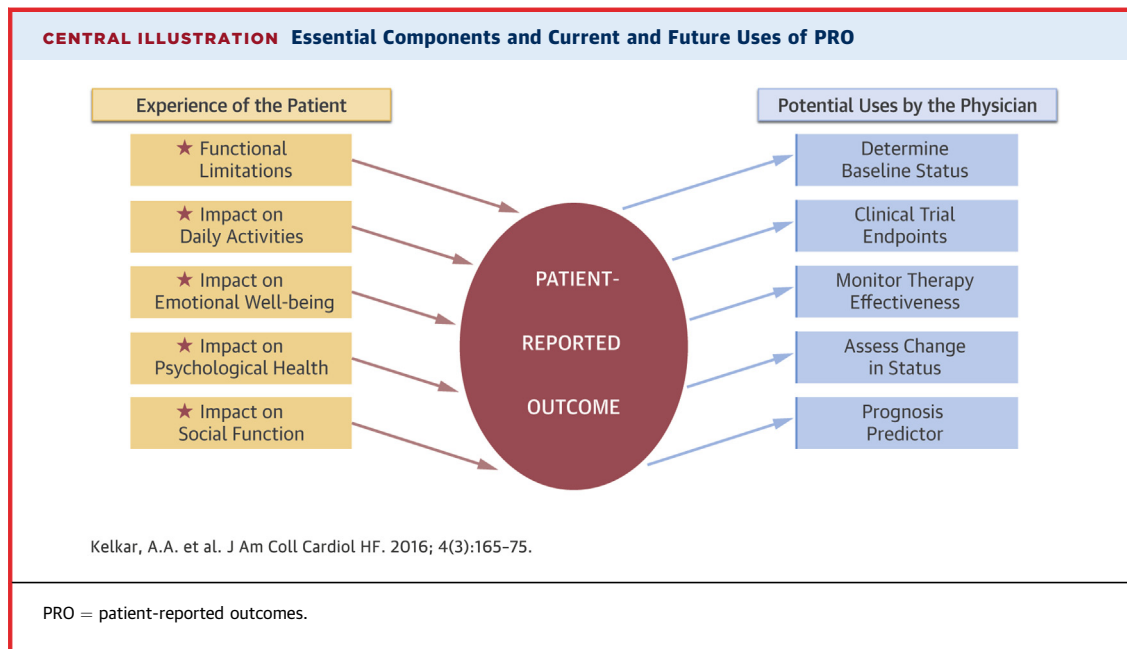
Symptom coverage. There is an inherent tension between creating and using tools that cover all potential manifestations of HF and are sensitive to patients' response burden in completing the instrument. **Table 2** describes 30 potential HF symptom and impact endpoints cross-matched with each PRO instrument. The KCCQ (14) and MLHFQ (13) had the highest coverage of these manifestations of HF. Both had the greatest emphasis on functional limitations

and HF-specific symptoms, but they also addressed emotional and relationship components. Conversely, the HFFSI and QLQ-SHF had the least amount of symptom coverage. No instruments addressed sudden weight gain or ascites.

DISCUSSION. Moving PRO from research tools to clinical care requires unique considerations, beyond the standard psychometric properties to emphasize feasibility and interpretability (14). The 2 instruments that best fit all of the evaluation criteria and meet the most symptom endpoints are the KCCQ and MLHFQ. These instruments are not only the most commonly used, but they also have been highly rated in systematic reviews (5). Although both of these instruments were developed with input from HF population, none of the instruments fully assessed all of the QoL domains that the HF patient experiences, including physical and mental/emotional symptoms, physical limitations, and social limitations. For example, both MLHFQ and KCCQ have a limited focus on emotional and relationship aspects (KCCQ and MLHFQ meeting 5 of 9 emotional endpoints). Although these measures have good psychometric properties, their use in routine clinical care and the impact of routine PRO use on treatment and outcomes has not been established.

Ultimately the KCCQ and MLHFQ instruments are both feasible to implement and have interpretable scores, but whether clinicians or patients understand these domain/scale scores or find them useful in the clinical setting has not been well described. Accordingly, there have been no studies demonstrating improvements in care with the use of these PRO. Even though this review suggests that either instrument might work well in the clinical setting, further research is needed to show that the tools are feasible and helpful in providing HF care, particularly to help prognosticate and risk stratify individual patients so that treatment can be more efficiently tailored to patients' risks. Whereas the brevity of PRO is important to support their feasibility in routine clinical use, whether these measures could be improved on by more focus on emotional and relationship aspects of the HF experience, or whether additional instruments that directly measure these domains should be used to supplement disease-specific PRO is an important avenue for future research.

A pragmatic approach to the use of PRO instruments is necessary. A large breadth of data exists highlighting the value of PRO, however their use in the clinical setting continues to lag. Clinical settings are limited by time and resource constraints. Questionnaires that are easy to administer and interpret are necessary. Most instruments required <10 min to

CENTRAL ILLUSTRATION Essential Components and Current and Future Uses of PRO

Experience of the Patient

★ Functional Limitations
★ Impact on Daily Activities
★ Impact on Emotional Well-being
★ Impact on Psychological Health
★ Impact on Social Function

PATIENT-REPORTED OUTCOME

Potential Uses by the Physician

Determine Baseline Status
Clinical Trial Endpoints
Monitor Therapy Effectiveness
Assess Change in Status
Prognosis Predictor

Kelkar, A.A. et al. J Am Coll Cardiol HF. 2016; 4(3):165–75.

PRO = patient-reported outcomes.

complete and the KCCQ-12 requires even less time. Thus the integration of these tools into clinical care may be more an issue of commitment than of patient-level burden.

All 9 PRO instruments studied differed not only in calculations but also in the interpretability of scores. Although some did not correlate to clinically meaningful differences, other instruments such as the KCCQ could clearly define whether a patient had had a large, moderate, or small improvement; this is a very important attribute if the instrument is to be a valuable component of clinical practice. Also, a PRO built for clinical use should be facile when administered or when scores are calculated and interpreted, and more research is needed to better define how to present cross-sectional and longitudinal changes in PRO scores to maximize their interpretation and usefulness. In a fashion similar to the education required for understanding what each NYHA functional class indicates, education as to what calculated PRO scores means is needed to aid their overall interpretability and usefulness in clinical care.

The clinical utility of PRO is a key feature of an ideal instrument. Studies have shown that physicians may underestimate the functional limitations of patients, underscoring the need for objective patient assessment [11]. PRO instruments objectively quantify patients' health status so that changes between current and previous health can be calculated and interpreted. The change in score may indicate the need to adjust a patient's medication; determine their

symptoms have worsened; or to involve therapists, counselors, or palliative care if QoL has deteriorated in the absence of a change in symptoms or functional limitations. Because PRO instruments extend beyond functional status and highlight other aspects of HF experience (loneliness, depression), use of these instruments potentially provides health care providers an easy way to readily assess broader impact of HF on patients' health status (**Central Illustration**). These tools could also potentially assist with patient decision making, as certain strategies may not necessarily improve longevity but can improve QoL (e.g., diuretic adjustments).

Future directions. A PRO built for clinical use should also be facile when administered or when scores are calculated and interpreted. More research is needed to better define how to present cross-sectional and longitudinal changes in PRO scores to maximize their interpretation and usefulness. However, given the importance that patients place on their health status (their symptoms, function, and QoL), PRO instruments may provide an important, systematic approach of obtaining QoL data that cannot be obtained from other measures such laboratory data or imaging [26–29]. Furthermore, PRO are being increasingly used in clinical trials as endpoints, and greater familiarity by clinicians in the use of these scores can accelerate their understanding and translation of clinical trial reports [19]. There are other potential uses, particularly in recently discharged patients, in whom scores below a critical

threshold may warrant closer follow-up. Achieving this potential requires clear evidence of the value of routine PRO use in clinical care.

To that end, we believe that it is time to conduct a cluster-randomized clinical trial where some practices routinely incorporate PRO into their care and are compared with usual care that does not use these measures in practice. Such a trial could then compare the outcomes (i.e., mortality, rehospitalization, and the health status of their patients) between clinics with and without access to routine use of PRO. Such a trial should likely focus on the use of PRO meeting the most psychometric criteria and that are deemed clinically feasible. Ultimately, the KCCQ and MLHFQ have the strongest overall evidence for use as a PRO instrument and should be explicitly tested to define whether they can improve the efficiency, patient-centeredness and quality of HF care. For this to occur, however, increasing education to providers on the value and interpretability of PRO instruments, as well as a more streamlined approach to their implementation in the clinical setting, are necessary.

REPRINT REQUESTS AND CORRESPONDENCE: Dr. Javed Butler, Cardiology Division, Stony Brook University, Health Sciences Center, T-16, Room 080, SUNY at Stony Brook, Stony Brook, New York 11794. E-mail: javed.butler@stonybrookmedicine.edu.

REFERENCES

1. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the Twenty-First Century. Washington, DC: National Academy Press, 2001.

2. Arnold SV, Spertus JA, Lei Y, et al. How to define a poor outcome after transcatheter aortic valve replacement: conceptual framework and empirical observations from the placement of aortic transcatheter valve (PARTNER) trial. Circ Cardiovasc Qual Outcomes 2013;6:591–7.

3. Spertus J, Peterson E, Conard MW, et al., for the Cardiovascular Outcomes Research Consortium. Monitoring clinical changes in patients with heart failure: a comparison of methods. Am Heart J 2005;150:707–15.

4. Mackintosh AG, Fitzpatrick R. A Structured Review of Patient-Reported Outcome Measures (PROMs) for Heart Failure. Oxford, UK: University of Oxford, 2009.

5. Garin O, Herdman M, Vilagut G, et al. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. Heart Fail Rev 2014;19: 359–67.

6. Centers for Medicare and Medicaid Services. 2015 National Impact Assessment of the Centers for Medicare and Medicaid Services (CMS) Quality Measures Report. Baltimore, MD: CMS, 2015.

7. Sepehrvand N, Ezekowitz JA. How to do more with less. Circ Cardiovasc Qual Outcomes 2015 Aug 25 [E-pub ahead of print].

8. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). London, UK: The Cochrane Collaboration, 2011.

9. Raphael C, Briscoe C, Davies J, et al. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart 2007;93: 476–82.

10. Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new specific activity scale. Circulation 1981;64:1227–34.

11. Calkins DR, Rubenstein LV, Cleary PD, et al. Failure of physicians to recognize functional disability in ambulatory patients. Ann Intern Med 1991;114:451–4.

12. Rector TS, Cohn JN, for the Pimobendan Multicenter Research Group. 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; 124:1017–25.

13. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. Circ Cardiovasc Qual Outcomes 2015 Aug 25 [E-pub ahead of print].

14. Spertus J. Barriers to the use of patient-reported outcomes in clinical care. Circ Cardiovasc Qual Outcomes 2014;7:2–4.

15. Anker SD, Agewall S, Borggrefe M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J 2014;35: 2001–9.

16. Feinstein AR. An additional basic science for clinical medicine: IV. The development of clinimetrics. Ann Intern Med 1983;99:843–8.

17. U.S. Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Washington, D.C.: U.S. Department of Health and Human Services, Food and Drug Administration, 2009.

18. McMurray JJ, Packer M, Desai AS, et al., for the PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371: 993–1004.

19. Rahimi K, Malhotra A, Banning AP, Jenkinson C. Outcome selection and role of patient reported outcomes in contemporary cardiovascular trials: systematic review. BMJ 2010;341:c5707.

20. Arnold SV, Spertus JA, Lei Y, et al. Use of the Kansas City Cardiomyopathy Questionnaire for monitoring health status in patients with aortic stenosis. Circ Heart Fail 2013;6:61–7.

21. Joseph SM, Novak E, Arnold SV, et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. Circ Heart Fail 2013;6:1139–46.

22. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol 2000;35:1245–55.

23. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. Circ Cardiovasc Qual Outcomes 2015 Aug 25 [Epub ahead of print].

24. Kosiborod M, Soto GE, Jones PG, et al. Identifying heart failure patients at high risk for near-term cardiovascular events with serial health status assessments. Circulation 2007;115:1975–81.

25. Riegel B, Moser DK, Glaser D, et al. The Minnesota Living With Heart Failure Questionnaire: sensitivity to differences and responsiveness to intervention intensity in a clinical population. Nurs Res 2002;51:209–18.

26. Heo S, Lennie TA, Okoli C, Moser DK. Quality of life in patients with heart failure: ask the patients. Heart Lung 2009;38:100–8.

27. Jiang W, Alexander J, Christopher E, et al. Relationship of depression to increased risk of mortality and rehospitalization in patients with congestive heart failure. Arch Intern Med 2001; 161:1849–56.

28. Faller H, Steinbuchel T, Störk S, Schowalter M, Ertl G, Angermann CE. Impact of depression on quality of life assessment in heart failure. Int J Cardiol 2010;142:133–7.

29. Juenger J, Schellberg D, Kraemer S, et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. Heart 2002;87:235–41.

30. Moher D, Liberati A, Tetzlaff J, et al., for the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6:e1000097. http://dx.doi.org/10.1371/journal.pmed1000097.

31. 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.

32. Bennett SJ, Perkins SM, Lane KA, Deer M, Brater DC, Murray MD. Social support and health-related quality of life in chronic heart failure patients. Qual Life Res 2001;10:671-82.

33. Mansfield DR, Gollogly NC, Kaye DM, Richardson M, Bergin P, Naughton MT. Controlled trial of continuous positive airway pressure in obstructive sleep apnea and heart failure. Am J Respir Crit Care Med 2004;169:361-6.

34. Guyatt GH, Sullivan MJ, Fallen EL, et al. A controlled trial of digoxin in congestive heart failure. Am J Cardiol 1988;61:371-5.

35. Spertus JA, Jones PG, Kim J, Globe D. Validity, reliability, and responsiveness of the Kansas City Cardiomyopathy Questionnaire in anemic heart failure patients. Qual Life Res 2008;17:291-8.

36. Kosiborod M, Krumholz HM, Jones PG, Pitt B, Spertus JA. The relationship between anemia, change in hematocrit over time and change in health status in patients with heart failure after myocardial infarction. J Card Fail 2008;14:27-34.

37. Conard MW, Heidenreich P, Rumsfeld JS, et al., for the Cardiovascular Outcomes Research Consortium. Patient-reported economic burden and the health status of heart failure patients. J Card Fail 2006;12:369-74.

38. Heidenreich PA, Spertus JA, Jones PG, et al., for the Cardiovascular Outcomes Research Consortium. Health status identifies heart failure outpatients at risk for hospitalization or death. J Am Coll Cardiol 2006;47:752-6.

39. Heo S, Doering LV, Widener J, Moser DK. Predictors and effect of physical symptom status on health-related quality of life in patients with heart failure. Am J Crit Care 2008;17:124-32.

40. Hou N, Chui MA, Eckert GJ, Oldridge NB, Murray MD, Bennett SJ. Relationship of age and sex to health-related quality of life in patients with heart failure. Am J Crit Care 2004; 13:153-61.

41. Bennett SJ, Oldridge NB, Eckert GJ, et al. Comparison of quality of life measures in heart failure. Nurs Res 2003;52:207-16.

42. Aranda JM Jr., Conti JB, Johnson JW, Petersen-Stejskal S, Curtis AB. Cardiac resynchronization therapy in patients with heart failure and conduction abnormalities other than left bundle-branch block: analysis of the Multicenter InSync Randomized Clinical Evaluation (MIRACLE). Clin Cardiol 2004;27:678-82.

43. Carels RA. The association between disease severity, functional status, depression and daily quality of life in congestive heart failure patients. Qual Life Res 2004;13:63-72.

44. Rector TS, Tschumperlin LK, Kubo SH, et al. Use of the Living With Heart Failure questionnaire to ascertain patients' perspectives on improvement in quality of life versus risk of drug-induced death. J Card Fail 1995;1:201-6.

45. Kato N, Kinugawa K, Seki S, et al. Quality of life as an independent predictor for cardiac events and death in patients with heart failure. Circ J 2011;75:1661-9.

46. Dunderdale K, Thompson DR, Beer SF, Furze G, Miles JN. Development and validation of a patient-centered health-related quality-of-life measure: the chronic heart failure assessment tool. J Cardiovasc Nurs 2008;23:364-70.

47. Shabetai R. Cardiomyopathy: how far have we come in 25 years, how far yet to go? J Am Coll Cardiol 1983;1:252-63.

48. Tate CW 3rd, Robertson AD, Zolty R, et al. Quality of life and prognosis in heart failure: results of the Beta-Blocker Evaluation of Survival Trial (BEST). J Card Fail 2007;13:732-7.

49. Dracup K, Walden JA, Stevenson LW, Brecht ML. Quality of life in patients with advanced heart failure. J Heart Lung Transplant 1992;11:273-9.

50. Dunbar SB, Jenkins LS, Hawthorne M, et al. Factors associated with outcomes 3 months after implantable cardioverter defibrillator insertion. Heart Lung 1999;28:303-15.

51. Stephen SA. Fatigue in older adults with stable heart failure. Heart Lung 2008;37:122-31.

52. Wiklund I, Lindvall K, Swedberg K, Zupkis RV. Self-assessment of quality of life in severe heart failure: an instrument for clinical use. Scand J Psychol 1987;28:220-5.

53. Wiklund I, Waagstein F, Swedberg K, Hjalmarsson A. Quality of life on treatment with metoprolol in dilated cardiomyopathy: results from the MDC trial. Metoprolol in Dilated Cardiomyopathy trial. Cardiovasc Drugs Ther 1996;10:361-8.

54. O'Leary CJ, Jones PW. The left ventricular dysfunction questionnaire (LVD-36): reliability, validity, and responsiveness. Heart 2000;83: 634-40.

55. Conti JB, Sears SF. Cardiac resynchronization therapy: can we make our heart failure patients smarter? Trans Am Clin Climatol Assoc 2007;118: 153-64.

56. Zambroski CH, Lennie TA, Chung ML, Heo S, Smoot T, Ziegler C. Use of the Memorial Symptom Assessment Scale-Heart Failure in heart failure patients. Circulation 2004;110 Supplement III:17.

57. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 1994;30A:1326-36.

58. Zambroski CH, Moser DK, Bhat G, Ziegler C. Impact of symptom prevalence and symptom burden on quality of life in patients with heart failure. Eur J Cardiovasc Nurs 2005;4:198-206.

59. Aiken LS, Butner J, Lockhart CA, Volk-Craft BE, Hamilton G, Williams FG. Outcome evaluation of a randomized trial of the PhoenixCare intervention: program of case management and coordinated care for the seriously chronically ill. J Palliat Med 2006;9:111-26.

KEY WORDS heart failure, patient-reported outcomes

APPENDIX For supplemental material, please see the online version of this article.