



HHS Public Access

Author manuscript

Circ Heart Fail. Author manuscript; available in PMC 2020 June 05.

Published in final edited form as:

Circ Heart Fail. 2019 June ; 12(6): e005751. doi:10.1161/CIRCHEARTFAILURE.118.005751.

The Development and Initial Validation of the PROMIS-Plus-HF Profile Measure

Faraz S. Ahmad, MD, MS^{a,b}, Michael A. Kallen, PhD, MPH^c, Karen E. Schifferdecker, PhD, MPH^{d,e,f}, Kathleen L. Carluzzo, MS^{e,f}, Susan E. Yount, PhD^c, Jill M. Gelow, MD, MPH^{g,h}, Peter A. McCullough, MD, MPHⁱ, Stephen E. Kimmel, MD, MSCE^j, Elliot S. Fisher, MD, MPH^{d,f}, David Cella, PhD^{c,k}

^aDivision of Cardiology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

^bThe Center for Health Information Partnerships (CHiP), Institute of Public Health & Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

^cDepartment of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois

^dCommunity and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

^eCenter for Program Design and Evaluation at Dartmouth, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

^fThe Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

^gOregon Health & Science University Knight Cardiovascular Institute, Portland, Oregon

^hProvidence Heart and Vascular Institute, Portland, Oregon

ⁱBaylor University Medical Center, Baylor Heart and Vascular Institute, Baylor Jack and Jane Hamilton Heart and Vascular Hospital, Dallas, Texas

^jDepartments of Medicine and of Biostatistics, Epidemiology & Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

^kCenter for Patient Centered Outcomes, Institute of Public Health & Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

Abstract

Background: Bringing together generic and heart failure (HF)-specific items in a publicly-available, patient-reported outcome measure may facilitate routine health status assessment for improving clinical care and shared decision-making, assessing quality of care, evaluating new interventions, and comparing groups with different conditions.

Address for Correspondence: Faraz Ahmad, MD MS. 676 North Saint Clair Street, Suite 600, Chicago, IL 60611. Office: 312-695-4965. Fax: 312-926-8620. faraz.ahmad@northwestern.edu.

Disclosures: None

Methods and Results: We performed a mixed-methods study to develop and validate the PROMIS®-Plus-HF profile measure, a HF-specific instrument based on the generic The Patient-Reported Outcomes Measurement Information System (PROMIS). We conducted eight focus groups with 61 HF patients and phone interviews with 10 HF clinicians. The measure was developed via an iterative process of reviewing existing PROMIS items and developing and testing new HF items. In 600-patient sample, we estimated reliability (internal consistency; test-retest, with n=100 participants). We conducted validity analyses using Pearson *r* and Spearman *rho* correlations with Kansas City Cardiomyopathy Questionnaire (KCCQ) subscores. In a longitudinal sample, we performed responsiveness testing (paired t-tests) with 75 HF patients receiving interventions with expected health status improvement. The PROMIS-Plus-HF measure comprises 86 items (64 existing; 22 new) across 18 domains. Internal consistency reliability (Cronbach's alpha) coefficients ranged from 0.52–0.96, with alpha 0.70 in 12/17 domains. Test-retest intraclass correlation coefficients were 0.90. Correlations with KCCQ subscores supported expected convergent ($r/\rho > 0.60$) and divergent validity ($r/\rho < 0.30$). In the longitudinal sample, 10/18 domains had improved ($P < 0.05$) scores from baseline to follow-up.

Conclusions: The PROMIS-Plus-HF profile measure—a complete assessment of physical, mental, and social health—exhibited good psychometric characteristics and may facilitate patient-centered care and research. Subsets of domains and items can be used depending on the clinical or research purpose.

Keywords

Heart Failure; Quality and Outcomes

Introduction

The integration of patient perspectives on their health comprises a key component of high quality, patient-centered care. The quantification of health status has myriad applications, including for improving clinical care and shared decision-making, assessing quality of care, and evaluating new interventions. Yet, in practice, patient perspectives are rarely incorporated in a systematic and clinically meaningful way, especially in cardiovascular disease.¹

Reengineering the health system inclusive of patient perspectives is particularly important for heart failure (HF), a common, costly, and morbid condition affecting over 6.5 million US adults with a 5-year survival rate of approximately 50%.² Although the defining clinical symptoms of HF are shortness of breath, fatigue, and exercise intolerance, the experience of patients with HF extends well beyond these symptoms and includes a range of physical, mental, and social effects.

Patient-reported outcomes measures (PROMs) may be generic or disease-specific and provide standardized methods of quantifying health status for individuals. Generic PROMs, such as PROMIS® (The Patient-Reported Outcomes Measurement Information System®) and SF-36, capture physical, mental, and social health information independent of a specific disease.^{3–5} In contrast, disease-specific measures evaluate health status in the context of a specific illness and historically tended to be more responsive to change over time and more

easily interpretable by specialists. In HF research, the two most widely used PROMs are disease-specific measures: the Kansas City Cardiomyopathy Questionnaire (KCCQ) and the Minnesota Living with Heart Failure Questionnaire (MLHFQ).^{6, 7} In clinical practice, the clinician-assigned New York Heart Association (NYHA) Class represents the most commonly used HF health status measure. However, NYHA Class crudely measures only a portion of health status and represents the clinician's interpretation of patient symptoms. Neither the KCCQ nor the MLHFQ is used routinely in clinical practice.⁶ In addition, the KCCQ and MLHFQ do not capture overall physical, mental, and social health status, which is important in conditions such as HF where patients frequently experience multiple illnesses and co-morbidities. Because KCCQ and MLHFQ the items are specific to patients with HF, they are not as conducive to comparisons across populations with different conditions.

Recently, Schifferdecker et al.⁸, using knee osteoarthritis as an exemplar condition, described the process for creating condition-specific assessments within generic PROMs such as PROMIS, so as to capture "the health burden imposed by specific problems" and yet retain "the ability to compare across diseases, conditions, populations, and systems." In the case of HF, bringing together generic and HF-specific items into a hybrid, publicly-available instrument relevant to patients with HF may facilitate better health status comparisons with other conditions and the general population (i.e. those with and without HF), between HF subgroups (e.g., those with chronic lung diseases vs. those with diabetes), and within individuals over time using shared common items. In contrast to using multiple, proprietary generic and disease-specific measures, the use of existing PROMIS items as a base may be more conducive to adoption by health systems in routine clinical practice due to ongoing efforts by electronic health records vendors, health system administrators, and researchers to integrate PROMIS measures as part of routine care in electronic health records and patient portals.^{9, 10} The objective of this study was to develop a PROM for patients with HF that combines relevant, previously-tested generic PROMIS health measures with HF-specific item content.

Methods

Using a longitudinal transformation mixed-methods design,¹¹ we conducted a multiple-phase study to develop and evaluate the PROMIS-Plus-HF profile measure. This study was approved by the Internal Review Boards of Dartmouth College, Baylor Scott & White Research Institute, The University of Pennsylvania, Maine Medical Center, Mayo Clinic, Oregon Health and Science University, and Northwestern University. The other participating clinics and hospitals established IRB Authorization Agreements with Dartmouth College. All participants provided informed consent. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Development of the PROMIS-Plus-HF profile measure

The methods for developing the PROMIS-Plus-HF profile measure were similar to the development methods for the PROMIS-Plus measure for knee osteoarthritis; they have been previously described in detail, including participant recruitment, patient focus groups, clinician semi-structured interviews, item selection, drafting of new items and cognitive

testing, and finalizing the profile measure.⁸ Briefly, measure development was comprised of three parts. In part 1, across four sites, eight focus groups were conducted with groups of HF patients, and semi-structured phone interviews were conducted with clinicians (Table I in the online-only Data Supplement). From June 2014 to August 2014, focus groups and clinician interviews were performed by an experienced moderator (Interview Guides included in the online-only Data Supplement).

In part 2, all focus group and clinician interviews were transcribed and coded in Dedoose (Manhattan Beach, CA)¹² by two team members using a thematic analysis approach.¹³ A team of qualitative researchers then conducted a gap analysis to identify content themes consistent with existing PROMIS domains, themes not currently addressed by the existing domains (gap themes), and themes not relevant for further development, such as being relevant for only certain individuals due to their own unique conditions and circumstances. For themes that mapped to existing domains, the research team selected high quality, relevant PROMIS items from those domains. These items were selected based on a number of criteria including clarity, sex or opportunity bias, adequate specificity, and item psychometric quality.

All gap themes underwent further detailed review. For gap themes deemed relevant to HF, not covered by existing PROMIS items, and measurable by patient report, we drafted new items and mapped them to either existing PROMIS domains or categorized them under newly created domains, as warranted.

In part 3, all new gap items underwent cognitive testing, as guided by Willis.¹⁴ Two members of the study team (KS, KC) conducted phone interviews with a subset of 10 HF patients from the focus groups. In addition to audio recording each interview, participant responses were recorded on a standard form with notes regarding overall comprehension of each item and relevance to the participant. Team members discussed the results and decided whether to retain, revise, or drop each item. The final set of retained items underwent a translatability review in anticipation of future, multi-lingual translations.

Psychometric evaluation of the PROMIS-Plus-HF profile measure

Table 1 defines all of the measure testing categories and subcategories, along with the study population, methods of measurements, and statistical approaches we used. Additional testing categories that are important but not assessed directly in this study are: interpretability of scores, burden of respondent, and alternative modes and methods of administration. We assessed content validity during the profile measure development phase of the study. We evaluated all PROMIS-Plus-HF items and domains for reliability and construct and criterion-related validity with a cross-sectional sample and responsiveness with a longitudinal design.

Additional measures for validity testing

We used the KCCQ and PROMIS Global Health for additional validity testing. The KCCQ is a 23-item, self-administered measure that quantifies physical function, symptoms (frequency, severity and recent change), social function, self-efficacy and knowledge, and quality of life in patients with HF.¹⁵ The PROMIS Global Health measure is a generic

instrument that assesses an individual's physical, mental, and social health. This 10-item measure asks participants about their overall health (1 item), quality of life (1 item), physical health and physical functioning (2 items), social activities (2 items), mental health (2 items), fatigue (1 item), and pain (1 item).^{16–18} Two scores are produced: Physical Health and Mental Health.

Cross-sectional sample, including a subset of test-retest participants—

Participants in the cross-sectional sample were identified and recruited by the online panel company Opinions4Good (Op4G; Portsmouth, NH). Op4G emailed potential participants and determined eligibility based on responses to a series of screening questions (Table 2). We estimated a sample size of 600 participants would enable accurate and stable reliability testing and sufficient power for validity testing.¹⁹ To ensure a broad and diverse population, we specified quotas for age, sex, race/ethnicity, and self-reported functional status. Participants were recruited on a rolling basis until each quota was satisfied. Eligible respondents were included in the study and asked to complete the measures. We administered a survey retest to a sub-sample of 100 respondents three to seven days following their baseline survey administration. All participants were given a nominal fee for their participation. Responses were collected June to July 2015.

Longitudinal sample—Across eight health systems (Table I in the in the online-only Data Supplement), we identified participants via medical records with HF who met specified inclusion and exclusion criteria and fit into one of the following five treatment categories that would likely have improvement in HF-related symptoms and health status between baseline and follow-up three months later: 1) initiation of guideline-directed medical therapy after new HF diagnosis or first hospitalization; 2) cardiac rehabilitation for chronic stable HF; 3) initiation of cardiac resynchronization therapy; 4) implantation of left ventricular assist device; and 5) recent discharge after hospitalization primarily for HF (Table 2). In the fifth “recent HF discharge” group, to identify a select group of patients with a higher likelihood of improvement in functional status and a lower likelihood of adverse events, we excluded patients with a prior admission for HF within thirty days or if they had been discharged on home inotropes. These five treatment categories and the specific patient enrollment and follow-up times (Table II in the online-only Data Supplement) were selected by the research team, which included two HF clinicians.

We tailored recruitment protocols for each clinic and treatment pathway. Typically, recruitment involved an initial contact with eligible patients during a scheduled appointment or hospital stay, with follow-up by phone, email, or a mailed letter. Recruited participants were asked to complete the longitudinal survey at baseline and at three months after baseline. Patients were given two main options for completing electronic surveys at baseline and follow-up: at-home via an emailed link or in-clinic on an iPad provided by the study. We took several steps to reduce follow-up attrition, including increasing incentives from baseline (\$20) to follow-up (\$30), issuing a paper follow-up survey to accommodate HF patients without in-clinic follow-up appointments, and facilitating training and networking across research coordinators at participating sites to improve participant retention. We considered participation in the profile measure-testing to be complete when at least 80% of

the HF survey had been answered. The baseline and 3-month follow-up surveys were collected from July 2015 to March 2017.

Psychometric Analyses

Initial psychometric assessment: In the cross-sectional sample, we summarized results per measure (e.g., mean; standard deviation) and conducted classical item analyses (e.g., inter-item correlation, item-adjusted total score correlation).²⁰ We created raw summed scores for each domain and identified minimum and maximum possible scores. Raw summed score distributions of measures were graphically displayed to determine their nature (i.e., normal vs. skewed or having excess kurtosis). Internal consistency reliability was assessed using Cronbach's alpha, considering coefficients ≥ 0.70 to be adequate reliability for conducting group comparisons; and coefficients ≥ 0.90 as having adequate reliability for individual comparisons.²¹ Using a subset of $n=100$ cases from the cross-sectional data, we evaluated test-retest reliability using intraclass correlation coefficients (ICCs).

For all domains with at least four items, we conducted categorical confirmatory factor analysis (CCFA) to assess the dimensionality of each measure; we used polychoric correlations, the Mplus weighted least square mean-variance adjusted estimator, and cases without missing responses in our CCFA analyses. A single factor model was run per measure, and overall model fit was reviewed using published standards for excellent fit: Comparative Fit Index ≥ 0.95 , Tucker-Lewis Index ≥ 0.95 , root mean square error of approximation <0.06 , and weighted root mean residual <1.00).^{22, 23} For our Comparative Fit Index, Tucker-Lewis Index, and root mean square error of approximation fit indices, we used each CCFA model's reported scaled Satorra-Bentler chi-square value in our fit index estimation.

Using modern measurement theory approaches, we evaluated item response theory (IRT) modeling assumptions (e.g., local independence, monotonicity, item fit) and assessed differential item functioning (DIF) by select factors.^{24–26} DIF analyses investigate whether item performance is impacted by subgroup membership status (e.g., do males and females of equivalent domain health status respond differently to items due solely to their gender status?). Because of the need to incorporate new items into existing measures that had been developed from IRT graded response model estimation, we employed the graded response model for all IRT-based analyses. On domains with ≥ 4 items, DIF was assessed for sex, age, and education level, where sufficient subgroup sample sizes (minimum $n=200$) existed. DIF score impact was studied using unadjusted vs. DIF-adjusted theta estimates²⁵ derived from lordif analysis,²⁷ comparing theta differences by median standard error and effect size criteria.^{25, 28}

Construct and known-groups validity: Because there are no gold standard measures that can serve as validity measures for each of the domains in the PROMIS-Plus-HF profile measure, we used a combination of subscale scores from the KCCQ and scores from PROMIS Global Health Physical and Mental as comparators. In the cross-sectional sample, we conducted convergent validity analyses using Pearson r and Spearman ρ correlations (to account for skewed score distributions) with PROMIS-Plus-HF domains (Physical

function; Symptoms; Life satisfaction; Satisfaction with social roles and activities; Ability to engage in social roles and activities) and KCCQ subscores (KCCQ: Physical Limitation, KCCQ: Symptom Severity; KCCQ: Quality of Life; and KCCQ: Social Limitation) and defined convergent validity as r or $\rho > 0.60$.^{15, 29, 30} We conducted divergent validity analyses by comparing PROMIS-Plus-HF domains (Physical function; Symptoms; Satisfaction with social roles and activities; Ability to engage in social roles and activities) with KCCQ: self-efficacy and defined divergent validity as r or $\rho < 0.30$.

We then conducted known-groups validity testing by using analysis of variance (ANOVA) to compare PROMIS-Plus-HF profile measure scores of patients with (a) low PROMIS Global Health Physical scores (i.e., lowest tertile raw scores, from 4 to 10) vs. high Global Health Physical scores (highest tertile raw scores, from 14 to 20) and (b) low Global Health Mental scores (lowest tertile raw scores, from 4 to 10) vs. high Global Health Mental scores (highest tertile raw scores, from 14 to 20).¹⁶⁻¹⁸

Confirmatory Validation Testing: Using a similar approach, we conducted additional psychometric assessment and construct and known-groups validity testing in the baseline sample of participants enrolled the longitudinal sample. These analyses included measure summary statistics, internal consistency reliability testing, and construct and known-groups validity testing.

Responsiveness: We investigated within-person score changes from baseline to 3-month follow-up status using paired t-tests for each measure domain. We also compared responsiveness in the PROMIS-Plus-HF domains (Physical function; Symptoms; Life satisfaction; Satisfaction with social roles and activities; Ability to engage in social roles and activities) with similar KCCQ subscores (KCCQ: Physical Limitation, KCCQ: Symptom Severity; KCCQ: Quality of Life; and KCCQ: Social Limitation).

Handling missing data: Data were collected electronically to minimize missing data and allow for real-time tracking to identify potential issues. The survey was programmed to encourage, but not require, responses to all items. We employed a combination of strategies to address issues of missingness related to non-response. In general, for item-based analyses we used complete data and did not impute individual missing item responses. For score-based analyses, we computed total scores for established measures based on their existing scoring algorithms that accounted for possible missing item responses. For new measures, we computed total scores using simple proration, when a minimum of 50% of a domain's items had been completed.

Results

Development of the PROMIS-Plus-HF profile measure

We conducted eight focus groups with total of 61 patients with HF at the hospitals and clinics where they receive care and phone interviews with physicians (n=5) and nurse practitioners (n=5) specializing in HF from two sites (Dartmouth-Hitchcock Medical Center and Oregon Health & Science University Hospital) (Figure 1). Basic demographic information on the 61 patients are summarized in Table 3. Analyses of these interviews led

to the identification of 231 unique codes. No new domains or items were identified from analysis of the clinical interviews; thus, we included focus group data only in subsequent analyses. After removing codes not relevant for further development, we mapped 93 codes to existing PROMIS measures within 13 domains, and identified 38 gap items that required measure development and mapping to either new or existing domains. After the iterative process of item review, drafting, revision, and cognitive testing of new items with 10 patients, the final profile measure was comprised of 86 (64 existing; 22 new) items from 18 domains, as shown in Table 4. Each item, except for items in the Dyspnea domain, is on a 5-point Likert scale (0–5) with the domain score equal to the sum of all items within the domain. The Dyspnea is on a scale of 1 to 4 with 5 as “I did not do this in the past 7 days.” All of the items are shown in Table III in the online-only Data Supplement. Readability assessment using the Lexile Analyzer[®] indicated that new items were interpretable at a fourth grade level; note that existing PROMIS items aim to be at a sixth grade or lower reading level.^{31, 32} In addition, based on the actual time it took respondents in the cross-sectional sample to complete the survey and on conservative estimates for typical PROMIS instrument completion of six items in one minute, we estimate that the PROMIS-Plus-HF instrument in its entirety takes about 14 to 15 minutes to complete. However, the profile measure was designed to be a library of measures; thus, the completion time will depend on the number of items selected by the researcher or clinician.

Testing of the PROMIS-Plus-HF profile measure

Table 3 summarizes the demographics of the focus groups (N=61), cross-sectional sample (N=600), and the longitudinal sample (N=75). In the cross-sectional sample, the mean age was 54 years. Nearly half of the cross-sectional respondents were female (45%), and 19% were black. The final longitudinal sample was comprised of 75 participants who completed both the baseline and follow-up surveys. In the longitudinal sample the mean age was 58 years. Nearly half of longitudinal participants were female (47%), and 43% identified as black or African American. In Figure I in the online-only Data Supplement, we depict the participant flow in the longitudinal sample and reasons for attrition from the 195 patients who originally enrolled and completed the baseline survey to the 75 patients who completed the longitudinal follow-up survey.

Initial psychometric analyses—We calculated raw summed scores for each domain and determined measure means, standard deviations, medians, minimums, maximums, and score distribution skewness and kurtosis; results are presented per domain in Table IV in the online-only Data Supplement. The score distributions were approximately normal, although some evidence of slight skewness and kurtosis was observed. Internal consistency reliability (Cronbach’s alpha) ranged from 0.52 (Life Satisfaction) to 0.94 (Depression); 12 of 17 domains (Anger is a 1-item measure) had alphas \geq 0.70 (Table 5). The average inter-item correlation for the HF measures ranged from 0.36 (Life Satisfaction) to 0.71 (Depression). Measures with internal consistency $<$ 0.70 (Health Behavior Outcomes, Cognitive Abilities, Life Satisfaction, Independence, and Social Isolation) tended to be shorter (two or three items in length) and have lower average inter-item correlations (ranging from 0.36 to 0.52) than the other HF measures.

For the test-retest reliability analyses of a subset of 100 individuals from the cross-sectional sample, the ICC estimates (including both systematic and random error) for the HF measures also demonstrated excellent test-retest reliability: They ranged from 0.90 (Independence) to 0.99 (Anxiety, Satisfaction with social roles and activities), with no ICC estimate <0.90. The ICCs (systematic + random vs. random error only) exhibited little to no differences (Table V in the online-only Data Supplement).

We performed CCFA analyses on the nine domains with a minimum of four items (Table VI in the online-only Data Supplement). Six of nine domains demonstrated excellent Comparative Fit Index and Tucker-Lewis Index-assessed model fit, with index values ≥ 0.95 . Model fit results based on root mean square error of approximation and weighted root mean residual indices indicated either good or excellent model fit. Overall, model fit indices provided evidence supportive of essential unidimensionality. No significant violations of IRT assumptions (including local independence, monotonicity, and item fit) were observed (data not shown). We conducted DIF studies using three factors, based on the availability of sufficient subgroup sample sizes (minimum $n=200$): sex (male vs. female), age (≤ 55 vs. >55), and education level (completed college or not). In Stage 1 of the DIF studies, during which items were flagged for potential DIF, no items were identified involving the three studied DIF factors for the HF measures analyzed. Therefore, no DIF Stage 2 analyses (i.e., DIF score impact studies) were conducted (Table VII in the online-only Data Supplement).

Construct and known-groups validity: For the expected convergent validity analysis, the evidence supported the convergent validity (r or $\rho > 0.60$) of the PROMIS-Plus-HF Physical function measure with KCCQ Physical Limitation ($r/\rho=0.71/0.62$); the PROMIS-Plus-HF Symptom measure with KCCQ: Symptom Severity ($r/\rho=0.66/0.58$); and the PROMIS-Plus-HF Life satisfaction with KCCQ: Quality of Life ($r/\rho=0.62/0.56$). The KCCQ: Social Limitation measure was correlated with the PROMIS-Plus-HF Satisfaction with social roles and activities ($r/\rho=0.62/0.56$) and the Ability to engage in social roles and activities ($r/\rho=0.60/0.52$) measures. For the divergent validity analysis, the PROMIS-Plus-HF domains (Physical function; Symptoms; Life satisfaction; Satisfaction with social roles and activities; Ability to engage in social roles and activities) and the KCCQ: self-efficacy had r or $\rho < 0.3$.

For the Global Health Physical known-groups validity comparison, members of the high-tertile score group had statistically significantly better domain status scores for all HF domains measured except Health Behavior (Table VIII in the online-only Data Supplement). For the Global Health Mental high vs. low tertile score group comparison, members of the high-score group had statistically significantly better domain status scores for all HF health domains measured (Table IX in the online-only Data Supplement).

Confirmatory Psychometric Assessment and Validity Testing: In the 185 participants who remained enrolled in the longitudinal sample throughout the study, the measure summary statistics, the internal consistency reliability testing, and construct and known-groups validity testing were all overall similar as shown in Tables X-XV in the online-only Data Supplement. Internal consistency reliability (Cronbach's alpha) ranged from 0.62 (Symptoms) to 0.96 (Dyspnea); 13 of 17 domains had alphas ≥ 0.70 . The average inter-item

correlation for the HF measures ranged from 0.37 (Symptoms) to 0.79 (Pain Interference). Selected PROMIS-Plus-HF measure domains showed expected convergent and divergent validity with KCCQ subscales and known-groups validity with the PROMIS Global Health Physical and Mental measure.

Responsiveness: We conducted paired *t* tests of baseline vs. follow-up HF profile measure scores to obtain evidence of within-person change across time. For ten of the domains (Dyspnea, Fatigue, Health Behavior Outcomes, Physical Function, Sleep Disturbance, Anger, Anxiety, Cognitive Abilities, Life Satisfaction, Satisfaction with Social Roles and Activities), HF patients had statistically significantly better domain status scores at follow-up, compared to their baseline status scores (Table 6). When we compared the responsiveness of the PROMIS-Plus-HF domains (Physical function; Symptoms; Life satisfaction; Satisfaction with social roles and activities; Ability to engage in social roles and activities) with similar KCCQ subscores (KCCQ: Physical Limitation, KCCQ: Symptom Severity; KCCQ: Quality of Life; and KCCQ: Social Limitation), we found that three of the five PROMIS-Plus-HF domains (Physical function; Life satisfaction; and Satisfaction with social roles and activities) showed responsiveness similar to the relevant KCCQ subscores (KCCQ: Physical Limitation, KCCQ: Quality of Life; and KCCQ: Social Limitation).

Discussion

We developed and evaluated the PROMIS-Plus-HF profile measure, a publicly-available measure that combines health status items from PROMIS and newly developed items with content identified through qualitative research with patients and clinicians (Figure 2). Overall, the measure exhibited good psychometric characteristics during initial validation testing. Approximately 74% of item content already existed in PROMIS items. Nevertheless, we identified some important new content and developed items for it, using a mixed-methods approach. In total, the HF profile measure contains 86 items across 18 domains and provides a comprehensive assessment of physical, mental, and social health status for patients with HF. These 86 items therefore comprise a comprehensive set of generic (PROMIS-based) items combined with disease-specific (“Plus”) items, offering the potential for both cross-disease comparability and within-disease specificity. Importantly, the entire measure is not intended to be administered at one time; domains should be selected based upon the purpose and intended use of a given project.

In a sample of 600 patients with HF, our reliability testing demonstrated good to excellent internal consistency in most domains, and excellent test-retest reliability in all domains. Our strategy was to retain items deemed important by patients and stakeholders. If items did not entirely fit CCFA or IRT models evaluated, they were nevertheless considered for retention within a measure or as stand-alone items. Comparisons with KCCQ subdomains and the PROMIS Global Physical and Mental measures supported the validity of the PROMIS-Plus-HF profile measure. Confirmatory reliability and validity testing were repeated using the baseline data in the in-person recruited longitudinal cohort from 7 health systems. The findings were overall similar as in the 600-person online cross-sectional sample. In the analysis of longitudinal data of 75 participants, we observed responsiveness in a subset of

domains in a clinical population with expected health status change over time and some overlap in responsiveness in comparison to KCCQ subscales.

The PROMIS-Plus-HF profile measure adds to the existing, expansive library of PROMs for patients with HF.^{6, 7} The KCCQ and the MLHFQ comprise the two most commonly used PROMs in clinical studies, are highly rated for HF, and have approval by the U.S. Food and Drug Administration for use as a medical device development tool; however, neither of these measures provides a full assessment of mental and social health.⁶ One strength of the PROMIS-Plus-HF profile measure is that it is a complete assessment of physical, mental, and social health status, constructed on the foundation of the extensively tested PROMIS generic health measures. The use of PROMIS as a foundation may facilitate implementation in a health system as part of routine care and enable comparisons across groups with different conditions.

The PROMIS-Plus-HF measure is a relatively long survey with a large number of items and domains. Fortunately, the modular nature of the PROMIS system—which has been extensively tested, used frequently in research studies, implemented in a small number of health systems, and incorporated into selected electronic health record systems by vendors—allows one to select the clinically most relevant content for any given setting or context.^{3, 4, 10} Thus, the entire profile measure need not be administered in practice. This is a notable difference from the KCCQ and MLHFQ questionnaires. For example, with PROMIS-Plus-HF, if the measure is implemented within a health system, a minimum set of items could be kept constant, but then patients could choose additional domains or items to track based on their preferences. This would retain measures that are important for clinicians and acknowledge what is important to the patient with HF, at the same time allowing health systems to collect the minimum set across all patients for system-wide measurement and comparisons. Similar strategies could be used in research depending on the focus and research questions.

There are several limitations to this study. First, our focus group participants were not fully representative of the US general population's educational, racial, and ethnic diversity. However, the cross-sectional and longitudinal samples were more representative. Second, we were not able to account for non-respondents in our cross-sectional panel. Third, five domains with 2 to 3 items had a Cronbach's alpha <0.70, which limits reliability of longitudinal testing for these domains. Fourth, we were unable to retain the majority of respondents with HF in the longitudinal study although there was no difference in mean score between those who completed the study and those who dropped out and baseline characteristics were overall similar. Fifth, because only nine participants completed the measure via paper, we were unable to test if administration of the PROMIS-Plus-HF via paper has similar psychometric performance as with electronic administration. Lastly, for validity testing of physical function items, we used well-established PROMs, such as KCCQ and the PROMIS Global Health measure, instead of objective measures of functional status, such as six-minute walk and cardiopulmonary exercise testing, which were beyond the scope of this study. Future research will examine with correlation between the PROMIS-Plus-HF physical function items and objective clinical measures and perform responsiveness testing in a larger sample.

In summary, we describe the development and initial validation of the PROMIS-Plus-HF profile measure—a new, publicly-available PROM for patients with HF that combines previously-tested, generic PROMIS health measures with HF-specific items. This study represents the initial development and validation of this measure. In addition to further reliability, validity, and responsiveness testing, future measure development should test strategies to improve its usability by reducing the number of items required for typical administration while maintaining measurement strengths. These strategies include the development of computer-adaptive testing models, the creation of a short-form version, and the development of summary scores for physical, mental, and social health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We wish to thank our valuable research partners: Drs. Marie Bakitas, Douglas Sawyer, and the committed staff at our partner sites. We are especially grateful for the input of our patient and clinician participants and for the guidance and advice from our Patient and Family Advisory Committee: Annette Jo Giarrante, Janet Trzaska, David Swanz, Jeff Gardner, Carol DuBois, Roger Arend, and Linda Wilkinson. We would also like to thank Dr. Mitchell Psotka for his editorial feedback. The statements presented in this work are solely the responsibility of the author(s) and do not necessarily represent the official views of the Patient-Centered Outcomes Research Institute® (PCORI®), the PCORI® Board of Governors or Methodology Committee, the Agency for Healthcare Research and Quality, or the National Heart, Lung, and Blood Institute.

Funding Sources: Research reported in this work was partially funded through a Patient-Centered Outcomes Research Institute® (PCORI®) Award (ME-1303–5928). Dr. Ahmad was in part supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award number T32HL069771 and the Agency for Healthcare Research and Quality under Award K12HS026385.

References

1. Rumsfeld JS, Alexander KP, Goff DC, Graham MM, Ho PM, Masoudi FA, Moser DK, Roger VL, Slaughter MS, Smolderen KG, Spertus JA, Sullivan MD, Treat-Jacobson D, Zerwic JJ and American Heart Association Council on Quality of Care and Outcomes Research, Council on Cardiovascular and Stroke Nursing, Council on Epidemiology and Prevention, Council on Peripheral Vascular Disease, and Stroke Council. Cardiovascular health: the importance of measuring patient-reported health status: a scientific statement from the American Heart Association. *Circulation*. 2013;127:2233–2249. [PubMed: 23648778]
2. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, de Ferranti SD, Floyd J, Fornage M, Gillespie C, Isasi CR, Jimenez MC, Jordan LC, Judd SE, Lackland D, Lichtman JH, Lisabeth L, Liu S, Longenecker CT, Mackey RH, Matsushita K, Mozaffarian D, Mussolino ME, Nasir K, Neumar RW, Palaniappan L, Pandey DK, Thiagarajan RR, Reeves MJ, Ritchey M, Rodriguez CJ, Roth GA, Rosamond WD, Sasson C, Towfighi A, Tsao CW, Turner MB, Virani SS, Voeks JH, Willey JZ, Wilkins JT, Wu JH, Alger HM, Wong SS, Muntner P, American Heart Association Statistics C and Stroke Statistics S. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation*. 2017;135:e146–e603. [PubMed: 28122885]
3. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R and Group PC. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010;63:1179–94. [PubMed: 20685078]
4. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M and Group PC. The Patient-Reported Outcomes Measurement Information System (PROMIS):

progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. 2007;45:S3–S11.

5. Ware JE Jr. and Sherbourne. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–83. [PubMed: 1593914]
6. Kelkar AA, Spertus J, Pang P, Pierson RF, Cody RJ, Pina IL, Hernandez A and Butler J. Utility of Patient-Reported Outcome Instruments in Heart Failure. *JACC Heart Fail*. 2016;4:165–75. [PubMed: 26874386]
7. Thompson LE, Bekelman DB, Allen LA and Peterson PN. Patient-reported outcomes in heart failure: existing measures and future uses. *Curr Heart Fail Rep*. 2015;12:236–46. [PubMed: 25703897]
8. Schifferdecker KE, Yount SE, Kaiser K, Adachi-Mejia A, Cella D, Carluzzo KL, Eisenstein A, Kallen MA, Greene GJ, Eton DT and Fisher ES. A method to create a standardized generic and condition-specific patient-reported outcome measure for patient care and healthcare improvement. *Qual Life Res*. 2018;27:367–378. [PubMed: 28795261]
9. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, Pearman T, Gershon R, Penedo FJ, Rosen S and Cella D. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2015;121:927–34. [PubMed: 25376427]
10. Blumenthal KJ, Chang Y, Ferris TG, Spirt JC, Vogeli C, Wagle N and Metlay JP. Using a Self-Reported Global Health Measure to Identify Patients at High Risk for Future Healthcare Utilization. *J Gen Intern Med*. 2017;32:877–882. [PubMed: 28341894]
11. Schifferdecker KE and Reed VA. Using mixed methods research in medical education: basic guidelines for researchers. *Med Educ*. 2009;43:637–44. [PubMed: 19573186]
12. Dedoose [web application]. Version 7.0.23. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2016.
13. Boyatzis RE. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks: SAGE Publications; 1998.
14. Willis G, Reeve BB and Barofsky I. The use of cognitive interviewing techniques in quality of life and patient-reported outcomes assessment. *Outcomes Assessment in Cancer: Measures, Methods and Applications*. 2005:610–622.
15. Green CP, Porter CB, Bresnahan DR and 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. [PubMed: 10758967]
16. Hays RD, Bjorner JB, Revicki DA, Spritzer KL and Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*. 2009;18:873–80. [PubMed: 19543809]
17. Hays RD, Schalet BD, Spritzer KL and Cella D. Two-item PROMIS(R) global physical and mental health scales. *J Patient Rep Outcomes*. 2017;1:2. [PubMed: 29757325]
18. Hays RD, Spritzer KL, Thompson WW and Cella D. U.S. General Population Estimate for “Excellent” to “Poor” Self-Rated Health Item. *J Gen Intern Med*. 2015;30:1511–6. [PubMed: 25832617]
19. Reise SP and Yu J. Parameter Recovery in the Graded Response Model Using MULTILOG. *J Educ Meas*. 1990;27:133–144.
20. PROMIS Cooperative Group. PROMIS® instrument development and validation scientific standards version 2.0. HealthMeasures Website. http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf 2013:1–72. Published 2013. Accessed May 2, 2019.
21. Cronbach L *Essentials of Psychological Testing*. 2 ed. New York: Harper; 1960.
22. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D and Group PC. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45:S22–31. [PubMed: 17443115]
23. Cook KF, Kallen MA and Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT’s unidimensionality assumption. *Qual Life Res*. 2009;18:447–60. [PubMed: 19294529]

24. Ponocny I Nonparametric goodness-of-fit tests for the rasch model. *Psychometrika*. 2001;66:437–459.
25. Choi SW, Gibbons LE and Crane PK. lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *J Stat Softw*. 2011;39:1–30.
26. Cella M, Knibbe C, Danhof M and Della Pasqua O. What is the right dose for children? *Br J Clin Pharmacol*. 2010;70:597–603. [PubMed: 21087295]
27. Choi SW, Gibbons LE and Crane PK. lordif: logistic ordinal regression differential item functioning using IRT [computer program]: The Comprehensive R Archive Network; 2016.
28. Cohen J A power primer. *Psychological bulletin*. 1992;112:155–9. [PubMed: 19565683]
29. Spertus JA and Jones PG. Development and Validation of a Short Version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes*. 2015;8:469–76. [PubMed: 26307129]
30. Campbell DT and Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959;56:81–105. [PubMed: 13634291]
31. The Lexile Analyzer®. Durham, NC: Metametrics®; 2017.
32. Swartz CW, Burdick DS, Hanlon ST, Stenner AJ, Kyngdon A, Burdick H and Smith M. Toward a theory relating text complexity, reader ability, and reading comprehension. *J Appl Meas*. 2014;15:359–71. [PubMed: 25232670]

What is New?

- We developed and validated a novel, patient-reported outcome measure of physical, mental, and social health of patients with heart failure built upon The Patient-Reported Outcomes Measurement Information System (PROMIS), a validated set of person-centered measures used across the world.
- The PROMIS-Plus-HF profile measure is a library of measures and items; users can select subsets of domains and items and create customized short form versions based on the clinical need or research question.

What are the Clinical Implications?

- The PROMIS-Plus-HF profile measure will enable clinicians and researchers to obtain customized information on the physical, mental, social health of patients with heart failure.
- These data can be used for myriad purposes, including enhancing quality of care, comparing new interventions, and improving shared-decision making.

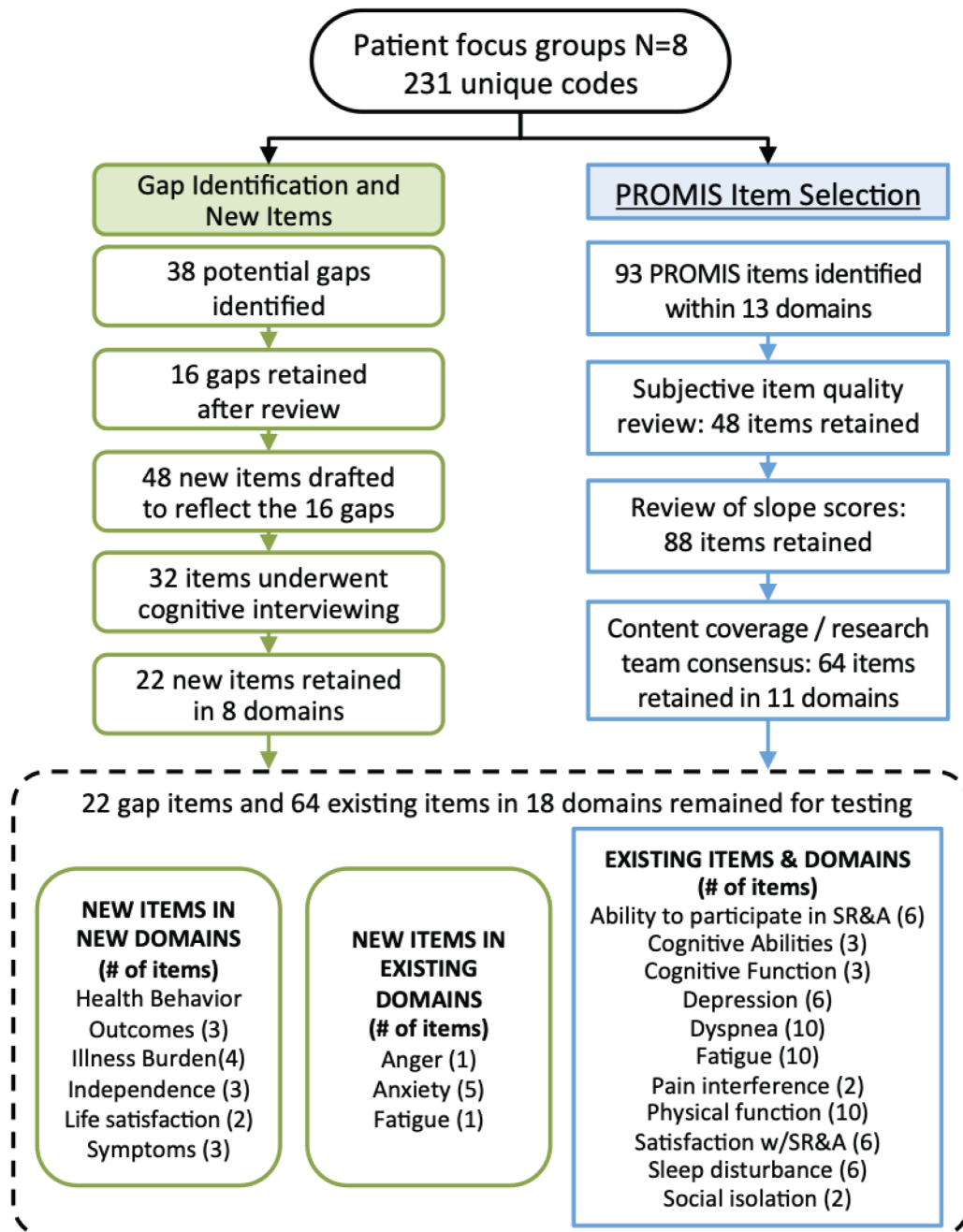


Figure 1. Gap analysis and item selection:

Analysis of interviews with 61 patients from 8 focus groups identified 231 unique codes. Through an iterative process, a total of 86 items across 18 domains were identified. Approximately 74% of item content already existed in PROMIS items. PROMIS = Patient-Reported Outcomes Measurement Information System. SR&A=Social Roles and Activities.

The PROMIS-Plus-HF Profile Measure

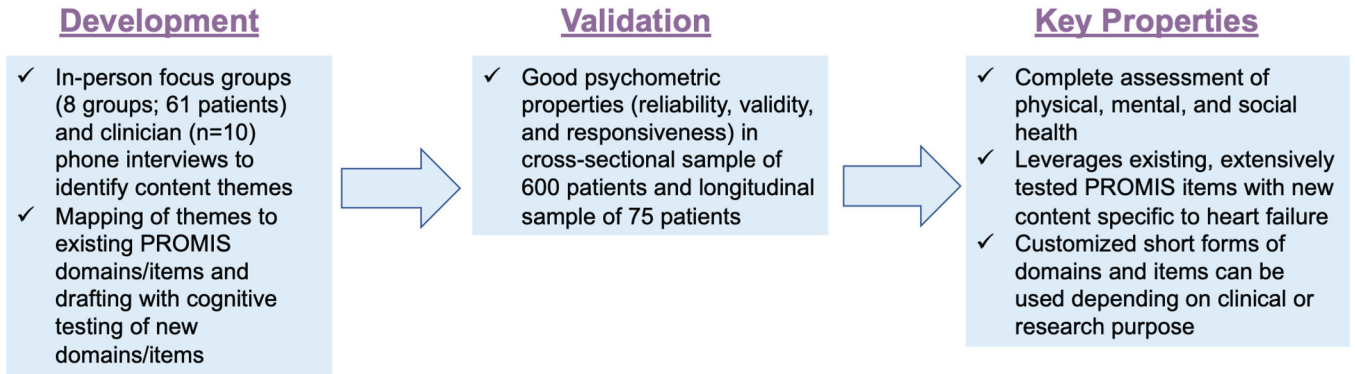


Figure 2. The PROMIS-Plus-HF profile measure: a complete assessment of physical, mental, and social health for patients with heart failure.

Table 1.

Key Categories and Approaches for Measure Testing

Category	Definition	Study population	Statistical approaches
Reliability			
Internal consistency	Extent of items measuring the same concept	Cross-sectional sample of 600 patients	Cronbach's alpha, inter-item correlation, item-adjusted total correlation, categorical confirmatory factor analysis, item response theory, differential item functioning
Test-retest reliability	Stability of responses over time without any clinical change	Repeat testing within 7 days for 100 participants	Intraclass correlation coefficient
Validity			
Content Validity	Instruments measures concept of interest	Qualitative focus groups with 61 patients, cognitive interviews with 10 patients, and interviews with 10 clinicians during measure development	n/a
Construct and Criterion-Related Validity	Data fit with prior, hypothesized relationships among items and domains and correlation with similar measure	Cross-sectional assessment of 600 patients with heart failure and correlation with KCCQ subdomains and known-groups validity comparisons with PROMIS Global Physical and Mental	Pearson and Spearman correlations, intraclass correlation coefficient, analysis of variance
Responsiveness	Measures show expected changes with interventions over time consistent with pre-determined hypotheses	Longitudinal sample of 75 patients with heart failure before and after one of five different interventions/clinical scenarios	Paired t-test

Adapted from: Cella D, Hahn EA, Jensen SE, et al. Patient-Reported Outcomes in Performance Measurement. September 2015. doi: <http://10.3768/rtipress.2015.bk.0014.1509>.

KCCQ = Kansas City Cardiomyopathy Questionnaire; PROMIS = Patient-Reported Outcomes Measurement Information System

Table 2.

Inclusion and exclusion criteria by target condition and sample

Cross-sectional sample*	Longitudinal sample*
<p>Inclusion:</p> <ul style="list-style-type: none"> • Diagnosis of HF • Age 18 years • Ability to speak English 	<p>Inclusion:</p> <ul style="list-style-type: none"> • Clinical diagnosis of HF documented in the medical record • Age 18 years • Ability to speak English • Part of one of the following groups: <ul style="list-style-type: none"> - Medication management (new onset or first hospitalization for heart failure) - Cardiac rehabilitation - Cardiac resynchronization therapy - Left ventricle assist device - Hospitalization primarily for HF
<p>Exclusion:</p> <ul style="list-style-type: none"> • Dementia • Serious mental disorder • Osteoarthritis of the Knee[‡] 	<p>Exclusion:</p> <ul style="list-style-type: none"> • Those scheduled for or have had a heart transplant. • Those with dementia or serious mental disorder. • Those in the "hospitalized primarily for HF" treatment group were excluded at baseline if they had a primary discharge diagnosis of HF within 30 days or at follow-up if discharged on home inotropes.
<p>Quotas:</p> <ul style="list-style-type: none"> • Age: 18–59 years (30%), 60 years (70%) • Sex: Female (50%), Male (50%) • Race: White (70%), Non-white (30%) • Severity of symptoms (based on NYHA Classification): • Class I: 10%, Class II-III: 80%, Class IV: 10% 	<p>Quotas: None</p>

* Eligibility data from the cross-sectional sample are self-reported; eligibility data from the longitudinal sample were drawn from medical records.

[‡] Patients with osteoarthritis were excluded to avoid the potential for overlap with a coincident study of testing patient-reported outcomes measures in this population
 HF = Heart Failure; NYHA = New York Heart Association

Table 3.

Characteristics of samples for measure development and testing

	Focus Group, N=61	Cross-sectional sample, N=600	Longitudinal sample, N=75
Age, years, mean (SD)	68 (13)	54 (14)	58 (12)
Sex, N (%)			
Female	24 (39)	270 (45)	35 (47)
Male	37 (61)	330 (55)	40 (53)
Race, N (%)			
American Indian or Alaska Native	1 (2)	17 (3)	1 (1)
Asian	0 (0)	39 (7)	1 (1)
Black or African American	4 (7)	115 (19)	32 (43)
Native Hawaiian or Pacific Islander	0 (0)	1 (<1)	0 (0)
White	55 (90)	401 (67)	38 (51)
Some other race	-	20 (3)	0 (0)
More than one race	0 (0)	7 (1)	0 (0)
Unknown or Not Reported	1 (2)	0 (0)	3 (4)
Ethnicity, N (%)			
Hispanic or Latino	1 (2)	171 (29)	2 (3)
Not Hispanic or Latino	60 (98)	429 (72)	73 (97)
Region of Enrollment, N (%)			
Midwest	32 (52)	-	14 (19)
Northeast	14 (23)	-	8 (11)
Pacific	15 (25)	-	15 (20)
South	0 (0)	-	38 (51)
Education Level, N (%)			
Did not complete high school	6 (10)	4 (0.7)	9 (12)
High school diploma or equivalent	6 (10)	73 (12)	23 (31)
Some college	19 (31)	154 (26)	21 (28)
Graduated college or higher	30 (49)	369 (61.5)	22 (29)
Categories, N (%)			
New diagnosis or first hospitalization	-	-	11 (15)
Cardiac rehabilitation	-	-	5 (7)
Cardiac resynchronization therapy	-	-	5 (7)
Left ventricular assist device	-	-	11 (15)
Hospitalization primarily for heart failure	-	-	43 (57)
Clinical Characteristics, N (%)			
Diabetes	-	80 (13)	28 (37)
Chronic obstructive lung disease	-	40 (7)	15 (20)
Depression	-	59 (10)	13 (17)
Chronic kidney disease	-	16 (3)	27 (36)
NYHA Class			

	Focus Group, N=61	Cross-sectional sample, N=600	Longitudinal sample, N=75
Class I	-	-	4 (5)
Class II	-	-	14 (19)
Class III	-	-	27 (36)
Class IV	-	-	6 (8)
Not available	-	-	24 (32)
Left ventricular ejection fraction < 50%	-	-	54 (72)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Summary of domains and items for PROMIS-Plus-HF profile measure

Domain Measure (n=18)	Domain includes new items	# Items Total	# Items Existing	# New Items	Range*
PHYSICAL					
Dyspnea	No	10	10	0	10–40
Fatigue	Yes	11	10	1	11–55
Health Behavior Outcomes	Yes	3	0	3	3–15
Pain Interference	No	2	2	0	2–10
Physical Function	No	10	10	0	10–50
Sleep Disturbance	No	6	6	0	6–30
Symptoms	Yes	3	0	3	3–15
MENTAL					
Anger	Yes	1	0	1	1–5
Anxiety	Yes	5	0	5	5–25
Cognitive Ability	No	3	3	0	3–15
Cognitive Function	No	3	3	0	3–15
Depression	No	6	6	0	6–30
Illness Burden	Yes	4	0	4	4–20
Life Satisfaction	Yes	2	0	2	2–10
SOCIAL					
Ability to Participate in Social Roles and Activities	No	6	6	0	6–30
Independence	Yes	3	0	3	3–15
Satisfaction with Social Roles and Activities	No	6	6	0	6–30
Social Isolation	No	2	2	0	2–10
TOTAL		86	64	22	

* Each item except for Dyspnea item is on a Likert scale ranging from 1 to 5. Dyspnea is on a scale of 1 to 4 with 5 as “I did not do this in the past 7 days.” The domain score is the sum score of the individual items.

Table 5.

Internal Consistency Reliability of the PROMIS-Plus-HF measure using cross-sectional sample data

Domain Measure	# items	alpha	Inter-item Correlation			Item-adjusted Total Correlation	
			Average	Min	Max	Min	Max
PHYSICAL							
Dyspnea	10	0.92	0.55	0.28	0.73	0.49	0.79
Fatigue	11	0.90	0.44	0.10	0.74	0.24	0.78
Health Behavior Outcomes	3	0.68	0.42	0.29	0.50	0.44	0.61
Pain Interference	2	0.76	0.62	0.62	0.62	0.62	0.62
Physical Function	10	0.90	0.49	0.06	0.68	0.37	0.77
Sleep Disturbance	6	0.88	0.55	0.41	0.73	0.61	0.77
Symptoms	3	0.72	0.47	0.43	0.53	0.50	0.58
MENTAL							
Anger	1						
Anxiety	5	0.86	0.55	0.51	0.61	0.64	0.71
Cognitive Abilities	3	0.64	0.37	0.17	0.59	0.29	0.62
Cognitive Function	3	0.87	0.70	0.69	0.71	0.75	0.76
Depression	6	0.94	0.71	0.64	0.78	0.75	0.85
Illness Burden	4	0.84	0.57	0.43	0.70	0.55	0.77
Life Satisfaction	2	0.52	0.36	0.36	0.36	0.36	0.36
SOCIAL							
Ability to Engage in Social Roles and Activities	6	0.90	0.59	0.46	0.72	0.58	0.78
Independence	3	0.66	0.40	0.26	0.58	0.35	0.58
Satisfaction with Social Roles and Activities	6	0.88	0.54	0.32	0.71	0.46	0.76
Social Isolation	2	0.68	0.52	0.52	0.52	0.52	0.52

Alpha = Cronbach's alpha

Paired t-test of PROMIS-Plus-HF domains and KCCQ subscales (Baseline vs. Follow-up) for the longitudinal sample

Table 6.

Domain Measure	Time	N	Mean	SD	t-test p value	Result Interpretation*
PHYSICAL						
Dyspnea	T1	55	24.53	9.16	0.004	T2 = better domain status
	T2	55	20.89	8.97		
Fatigue	T1	71	37.76	9.60	0.001	T2 = better domain status
	T2	71	34.21	10.59		
Health Behavior Outcomes	T1	75	10.81	2.78	0.002	T2 = better domain status
	T2	75	11.63	2.22		
Pain Interference	T1	74	5.30	2.45	0.453	No time difference
	T2	74	5.09	2.36		
Physical Function	T1	72	27.57	10.17	0.006	T2 = better domain status
	T2	72	30.25	9.94		
Sleep Disturbance	T1	74	19.69	6.67	0.002	T2 = better domain status
	T2	74	16.88	6.76		
Symptoms	T1	74	7.74	2.94	0.055	No time difference
	T2	74	7.12	2.76		
MENTAL						
Anger	T1	75	3.43	1.10	0.012	T2 = better domain status
	T2	75	3.04	1.37		
Anxiety	T1	72	13.58	4.51	0.032	T2 = better domain status
	T2	72	12.49	4.64		
Cognitive Abilities	T1	73	10.14	2.50	0.028	T2 = better domain status
	T2	73	10.93	3.08		
Cognitive Function	T1	74	10.88	3.33	0.492	No time difference
	T2	74	11.11	3.41		
Depression	T1	71	14.01	6.36	0.096	No time difference
	T2	71	13.03	6.23		
Illness Burden	T1	73	9.59	4.10	0.377	No time difference
	T2	73	10.03	4.35		

Domain Measure	Time	N	Mean	SD	t-test p value	Result Interpretation*
Life Satisfaction	T1	75	5.65	2.25	0.003	T2 = better domain status
	T2	75	6.35	2.23		
SOCIAL						
Ability to Engage in Social Roles and Activities	T1	72	15.86	6.02	0.084	No time difference
	T2	72	17.13	6.79		
Independence	T1	71	8.69	3.34	0.179	No time difference
	T2	71	9.25	3.54		
Satisfaction with Social Roles and Activities	T1	71	16.87	6.47	0.040	T2 = better domain status
	T2	71	18.37	6.61		
Social Isolation	T1	74	5.73	2.10	0.855	No time difference
	T2	74	5.77	2.08		
KCCQ SUBSCORES						
KCCQ: Physical Limitation	T1	71	46.98	27.10	0.002	T2 = better domain status
	T2	71	56.89	26.45		
KCCQ: Quality of Life	T1	73	36.42	26.26	<0.001	T2 = better domain status
	T2	73	54.00	28.26		
KCCQ: Self Efficacy	T1	71	77.91	26.15	0.013	T2 = better domain status
	T2	71	84.76	21.07		
KCCQ: Social Limitation	T1	63	40.67	31.11	<0.001	T2 = better domain status
	T2	63	54.27	32.95		
KCCQ: Symptom Frequency	T1	72	52.19	35.73	<0.001	T2 = better domain status
	T2	72	70.17	37.41		
KCCQ: Symptom Severity	T1	74	46.40	30.35	<0.001	T2 = better domain status
	T2	74	61.26	31.04		
KCCQ: Symptom Stability	T1	74	35.81	30.43	<0.001	T2 = better domain status
	T2	74	58.78	29.68		

* Interpretation of results of statistical F test

KCCQ = Kansas City Cardiomyopathy Questionnaire

SD = standard deviation