



OPEN ACCESS

EDITED BY

Alexander E. Berezin,
Paracelsus Medical University, Austria

REVIEWED BY

Attila Frigy,
George Emil Palade University of Medicine,
Pharmacy, Sciences and Technology of Târgu
Mureș, Romania
Natasa Sedlar,
General Hospital Murska Sobota, Slovenia

*CORRESPONDENCE

Istvan Mucsi
✉ Istvan.mucsi@utoronto.ca

[†]These authors have contributed equally to
this work

RECEIVED 02 April 2025

ACCEPTED 13 October 2025

PUBLISHED 30 October 2025

CITATION

Marblestone N, Chu S, Tomei N, Lodge D,
Bansal A, Edwards N, Ross HJ, Stehlik J,
Thayaparan D, Fadlallah J, Lee JG and Mucsi I
(2025) Validation of PROMIS anxiety item bank
computer adaptive test among patients with
heart failure.
Front. Cardiovasc. Med. 12:1605130.
doi: 10.3389/fcvm.2025.1605130

COPYRIGHT

© 2025 Marblestone, Chu, Tomei, Lodge,
Bansal, Edwards, Ross, Stehlik, Thayaparan,
Fadlallah, Lee and Mucsi. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Validation of PROMIS anxiety item bank computer adaptive test among patients with heart failure

Nolan Marblestone^{1†}, Steven Chu^{2,3†}, Nicole Tomei¹,
Denzel Lodge¹, Aarushi Bansal⁴, Nathaniel Edwards¹,
Heather J. Ross^{1,2,5}, Josef Stehlik⁶, Desana Thayaparan²,
Jad Fadlallah¹, Joshua G. Lee⁷ and Istvan Mucsi^{1*}

¹Ajmera Transplant Centre and Division of Nephrology, University Health Network, Toronto, ON, Canada, ²Ted Rogers Centre for Heart Research (TRCHR), Toronto, ON, Canada, ³ICES, Institute for Clinical Evaluative Services, Toronto, ON, Canada, ⁴Department of Medicine, Division of Cardiology, McMaster University, Hamilton, ON, Canada, ⁵Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada, ⁶Division of Cardiovascular Medicine, University of Utah, Salt Lake City, UT, United States, ⁷Department of Medical Sciences, Western University, London, ON, Canada

Introduction: Anxiety is highly prevalent among patients with heart failure (HF), negatively affecting health related quality of life (HRQOL). The Patient-Reported Outcomes Measurement Information System (PROMIS) anxiety item bank computer adaptive testing (CAT) precisely assesses anxiety symptom severity. This study aims to assess construct validity and reliability of PROMIS-Anxiety CAT among patients hospitalized for HF.

Methods: A cross-sectional convenience sample of adult patients hospitalized for HF, who completed PROMIS-A CAT, generalized anxiety disorder 7 (GAD-7), and other questionnaires electronically. Convergent validity was assessed by Spearman's rank correlation between PROMIS-A CAT, GAD-7, and other legacy measures. Known group analysis compared PROMIS-A CAT and GAD-7 scores between groups expected to have different levels of anxiety. Reliability of PROMIS-A CAT was calculated on the individual and group level from standard error of measurement, according to item response theory. Area under receiver-operating characteristics (ROC) curve and Youden's J statistic were used to identify a T-score cut-off for moderate/severe anxiety.

Results: Of 333 participants, 87 (26%) had moderate/severe anxiety based on GAD-7 score (≥ 10). Participants completed on average (median [IQR]) 4(1) vs. 7(0) items, with PROMIS-A CAT and GAD-7, respectively. PROMIS-A CAT T-scores were strongly correlated with GAD-7 scores ($\rho = 0.78$) and moderately correlated with other legacy measures. Known-group analysis provided further support for construct validity of PROMIS-A CAT. Individual reliability for PROMIS-A CAT T-scores was >0.9 for 87% of the sample; mean reliability was 0.91. Based on ROC and Youden's J analyses, a T-score of 60 can be used to identify individuals with moderate/severe anxiety.

Conclusion: These results support the validity and reliability of PROMIS-A CAT among patients hospitalized for HF.

KEYWORDS

PROMIS (Patient-Reported Outcomes Measurement Information System), PROMS (patient-reported outcome measures), anxiety, heart failure, validation, mental health

1 Introduction

Approximately 750,000 Canadians live with heart failure (HF), with about 100,000 new patients diagnosed annually (1). HF is characterized by signs and symptoms of congestion (i.e., shortness of breath, orthopnea, jugular venous distention, and pedal edema) that result from structural and/or functional cardiac abnormalities causing elevated cardiac pressures and reduced cardiac output (2). Patients with HF may experience diverse psychological and physical symptoms, which can contribute to impaired health-related quality of life (HRQOL) and increased healthcare utilization (3, 4).

Reportedly, 29%–53% of patients with HF have clinically relevant anxiety symptoms (3, 5–7). Anxiety is frequently underdiagnosed and undertreated in patients with chronic medical conditions, including HF (8, 9). Treating anxiety among patients with HF may improve outcomes, underscoring the importance of early screening, diagnosis, and treatment (8).

Patient-reported outcomes (PROs) are reports directly from patients regarding their functional abilities, symptoms, and feelings related to a health condition and its treatment (10). Patient-reported outcomes measures (PROMs) are standard questionnaires used to measure PROs, including anxiety (11). The use of PROMs when linked to appropriate symptom management pathways can improve clinical outcomes, quality of care, and communication between patients and healthcare providers (12–14).

The Generalized Anxiety Disorder 7 (GAD-7) questionnaire is a 7-item tool that is widely used to assess anxiety symptoms (15). However, tools like GAD-7 have been developed based on Classical Test Theory. These instruments include items that cover the whole symptom severity spectrum (including both the high and the low end), requiring all or most items to be completed by participants to obtain valid and reliable scores. Consequently, respondents may be obliged to complete irrelevant items, which can lead to high questionnaire burden, respondent fatigue, poor completion rates, and compromised data quality (16, 17).

The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and validated item banks to measure generic, clinically actionable PROs, relevant across various medical conditions (18). PROMIS item banks have been developed using the Item Response Theory (19, 20), where each item and every response option is calibrated to a *T*-score. Consequently, any combination from the item bank can be used, depending on the specific context. PROMIS and other tools developed by IRT can be administered as fixed-length short forms (SF); 2–4 item SF are typically the shortest options, with 4–10 item SF available as well. An alternative administration method is computer adaptive testing (CAT) (21). When administering an item bank via CAT, all participants answer an initial item that is calibrated for average symptom severity; subsequent items are selected by an algorithm based on prior responses, ensuring each question is relevant to an individual's symptom severity or level of functioning (22, 23). Items are delivered until a stopping rule is met; often when reliability over 90% is reached or 12 items are completed (24). CAT therefore delivers tailored questionnaires, omitting

irrelevant items while maintaining high precision, which may increase completion and adherence rates.

PROMIS instruments, therefore, offer excellent measurement precision with tailored questions to reduce question burden. Other studies have confirmed good measurement characteristics of the PROMIS-A CAT, reporting high reliability and good construct validity among multiple patient populations including those with chronic kidney disease and chronic pain (25, 26). Additionally, there is evidence supporting the feasibility of routine PROMIS-A CAT use in an in-patient setting (27). However, further analysis is required to ensure these measurement characteristics remain acceptable among diverse patient populations and to set population-specific, clinically actionable thresholds. This is the first study to assess the validity and reliability of PROMIS Anxiety Computer Adaptive Test (PROMIS-A CAT) in measuring anxiety symptoms among hospitalized patients with HF.

2 Methods

2.1 Study design & patient population

This analysis was completed with a cross-sectional, convenience sub-cohort of adult (≥ 18 years) patients experiencing HF, who were enrolled in the “*Predicting Readmission Outcomes using Biostatistical Evaluation and Machine Learning (PROBE ML)*” study at Toronto General Hospital and Toronto Western Hospital between March 2019 and October 2022. Patients were considered for inclusion in the study if they were admitted with a diagnosis of HF. Clinical diagnosis of HF was guided by the Framingham criteria for HF and/or serum BNP levels >100 pg/ml (28). Serum BNP was measured with a two-step chemiluminescent assay using the Abbott Architect i2000 analyzer. Patients were excluded (exclusion criteria were informed by the objectives of the *PROBE ML* study) if they were diagnosed with dementia, had severe cognitive deficits, underwent a heart or double lung transplant, had active cancer, did not speak English, were not Ontario residents, or were currently undergoing dialysis. Patients with a “Do Not Resuscitate” (DNR) order or those receiving end-of-life palliative care were also excluded. Similarly, patients who lived in long-term care or nursing home facilities, or were scheduled for discharge to such facility, were not enrolled as these patients were expected to require a different level of care and experience higher readmission rates (29). For this analysis, those who did not complete legacy questionnaires were also excluded. All participants provided written informed consent before enrolment. Research Ethics Board approval was obtained (REB#18-5658).

2.2 Questionnaire administration

Patients were approached by research team members during admission and invited to participate in the *PROBE ML* study.

Approximately the first 300 consenting participants were administered both PROMIS and legacy measures (established, valid questionnaires measuring the specific construct of interest) for validation of the PROMIS tools; this sample gives over 90% power to detect the target correlation ($\rho = 0.6$) at an alpha of 0.05. These participants completed multiple PROMIS domains assessed via CAT (PROMIS Bank v1.0: anxiety, fatigue, depression, dyspnea severity; PROMIS Bank v2.0: physical function, PROMIS Global Health 10 v1.1) and legacy questionnaires using a tablet-based electronic data capture (Data Driven Outcomes System, TECHNA Institute, University Health Network, Toronto).

2.3 Sociodemographic and clinical characteristics

Sociodemographic characteristics were obtained by trained chart abstractors from the Institute for Clinical Evaluative Sciences (ICES). Clinical laboratory variables were obtained from a blood sample (15 cc for plasma, serum, and buffy coat), collected within one day of study enrolment for the majority (90%) of participants. From these samples, hemoglobin, serum creatinine, serum sodium, and serum BNP level were measured. HF characteristics included etiology and HF type, classified as follows: HF with reduced ejection fraction (HFrEF) [left ventricular ejection fraction (LVEF) <40%]; HF with mildly reduced ejection fraction (HFmEF) (LVEF 40%–49%); or HF with preserved ejection fraction (HFpEF) (LVEF $\geq$ 50%). HF characteristics, etiology, and comorbidities [to calculate the Charlson Comorbidity Index (CCI)] were obtained from medical records.

2.4 PROMIS-anxiety item bank

The PROMIS Anxiety item bank v1.0 for adults includes 29 items that assess fear, anxious state, hyperarousal, and somatic experiences related to mental arousal. Each item requires patients to rate frequency of particular events on a 5-point Likert scale (“never”, “rarely”, “sometimes”, “often”, and “always”), where higher scores correspond to greater levels of anxiety (30). Raw scores are converted to a standardized *T*-score, where a mean score of 50 and standard deviation (SD) of 10 corresponds to the mean (SD) anxiety score of the United States general population (18). When administered via CAT, participants complete a minimum of 4 items. Items are administered until a standard error of measurement (SEM) of <0.3 (reliability >0.90) is achieved or participants have completed 12 items, according to the established stopping rule (24, 31, 32).

2.5 Legacy questionnaires

Construct validity of PROMIS-A CAT was assessed by analyzing correlations between PROMIS-A CAT *T*-scores and legacy questionnaire scores. The Generalized Anxiety Disorder-7 (GAD-7) was selected as the primary legacy instrument. This

7-item questionnaire assesses anxiety symptoms based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria for generalized anxiety (33). It uses a 4-point Likert scale ranging from 0 (“not at all bothered”) to 3 (“bothered nearly every day”) to measure severity of self-reported anxiety, with total scores ranging from 0 to 21. Scores $\geq$ 10 indicate moderate to severe anxiety symptoms (15).

The Edmonton Symptom Assessment System-revised (ESAS-r) measures the severity of nine emotional and physical symptoms on an 11-point scale, ranging from 0 (“no”) to 10 (“worst possible”) (34, 35). The anxiety item from this tool was used as the secondary legacy instrument for this analysis. The ESAS-r has demonstrated reliability and validity in patients with HF (36).

The EuroQol 5-Dimension 5-Level (EQ-5D-5L) assesses self-rated health over 5 domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (37, 38). The EQ-5D-Anxiety/Depression item was used as the tertiary legacy instrument for this analysis. It uses a 5-point scale, ranging from “not anxious or depressed” to “I am extremely anxious or depressed” (39). This questionnaire has documented reliability and validity among patients with heart disease (40).

The Patient Health Questionnaire-9 (PHQ-9) is a 9-item measure of depression, with each item assessed on a 4-point scale from “not at all” to “nearly every day”. Scores range from 0 to 27, with scores $\geq$ 10 representing moderate depression severity (41).

The Kansas City Cardiomyopathy Questionnaire-12 (KCCQ-12) is a 12-item, self-administered tool that assesses HRQOL in patients with HF. It consists of four domains: physical limitation, symptom frequency, quality of life, and social limitations; all are individually scored from 0 (worst possible health) to 100 (best possible health). The KCCQ-12 summary score is calculated as the mean of the four subdomain scores (42).

2.6 Statistical analysis

Baseline descriptive statistics are presented as mean (SD) for normally distributed variables, median [interquartile range (IQR)] for skewed variables, and frequency (%) for categorical variables. Characteristics between participants with vs. without moderate/severe anxiety (cut-off score: GAD-7 $\geq$ 10) were compared using independent sample *T*-tests for normally distributed variables and Mann-Whitney *U*-tests for nonparametric variables. Normality was assessed by a visual inspection of a density plot and QQ-plot for all variables, as well as Pearson’s coefficient of skewness for PROMs scores. Categorical variables were compared using chi-squared tests. Bonferroni correction for multiple tests was used to determine a significant alpha threshold by dividing an alpha of 0.05 by number of tests performed (43). Floor and ceiling effects were calculated as the percentage of participants who scored at the minimum and maximum possible questionnaire score, respectively. Skewness of PROM scores was quantified using Pearson’s moment coefficient of skewness. A coefficient of 0

represents a symmetric distribution (i.e., normal distribution); a large positive coefficient indicates a right-skewed distribution, whereas a large negative coefficient indicates left-skewness (44).

The reliability of the PROMIS-A CAT was assessed at both the individual and group level. We calculated individual-level reliability from SEMs across the PROMIS-A CAT score spectrum using the formula: $\text{reliability} = 1 - \text{SEM}^2$ to obtain values ranging from 0 (no reliability) to 1 (perfect reliability). Reliability ≥ 0.90 ($\text{SEM} = 0.32$), is considered acceptable for an individual score (45). Group level reliability was calculated using the formula: $\text{average reliability} = 1 - [\text{mean}(\text{SEM})]^2$, where reliability ≥ 0.90 is considered acceptable (31). Cronbach's alpha was calculated to assess internal consistency of the GAD-7, ESAS-r, and EQ-5D-5L. Alpha values between 0.80 and 0.89 indicate good internal consistency, while values >0.90 indicate excellent internal consistency (46).

Convergent validity was assessed by examining *T*-score correlations between PROMIS-A CAT and legacy measures assessing the same or similar construct (GAD-7, ESAS-r Anxiety, and EQ-5D-Anxiety/Depression). Moderate correlation ($\rho 0.5\text{--}0.7$) is considered acceptable, while a strong correlation ($\rho >0.7$) shows excellent validity (47, 48). Divergent (discriminant) validity was assessed by examining *T*-score correlations between PROMIS-A CAT and tools measuring constructs unrelated to anxiety (ESAS-r Appetite, KCCQ-Physical Limitation, and the EQ-5D-Mobility). We expected weak correlations ($\rho <0.4$) between these measures (49).

To further analyze construct validity, we compared mean PROMIS-A CAT *T*-scores and median GAD-7 scores between groups expected to have different levels of anxiety. We expected higher anxiety among participants who were female (50), were younger (51), had HF/rEF (52), and had greater comorbidity (53). We established additional groups based on legacy PROMs. We used the ESAS-r symptom score cut-off ≥ 30 to define high global symptom burden and expected these participants to have higher anxiety (23). We used a PHQ-9 score (cut-off ≥ 10) to define moderate to severe depressive symptoms and expected these participants to have higher anxiety (41, 54). Lastly, we used the KCCQ-12 summary score (cut-off ≥ 25) to indicate good HRQOL, expecting participants with impaired HRQOL (<25) to have higher average anxiety (55). For the known group analyses, PROMIS-A CAT *T*-scores and GAD-7 scores were adjusted for age, sex, and EF. Adjusted PROMIS-A CAT *T*-scores were obtained from linear regression model least-squared means. Adjusted median GAD-7 scores were obtained from quantile regression predicted medians. Group comparisons of adjusted mean PROMIS-A CAT *T*-scores and adjusted median GAD-7 scores were conducted using independent sample *T*-tests for binary group PROMIS-A CAT *T*-scores, Mann–Whitney *U*-tests for binary group GAD-7 scores, analysis of variance (ANOVA) for PROMIS-A CAT *T*-scores with more than 2 groups, and Kruskal–Wallis tests for GAD-7 scores with more than 2 groups. Cohen's *D* effect size was calculated for all binary comparisons of PROMIS-A CAT *T*-scores, which was classified as 0–0.49 (small), 0.5–0.79 (moderate), and >0.8 (large) (56). To account for the skewed distribution of

GAD-7 scores, bootstrap resampling with 1000 replication was used to estimate Cliff's delta effect size—classified as 0–0.32 (small), 0.33–0.46 (moderate) >0.47 (large) (57).

To assess discrimination of the PROMIS-A CAT, we conducted receiver operating characteristics (ROC) analysis using a GAD-7 score ≥ 10 (indicating moderate to severe anxiety) as the reference (15, 58). Test discrimination was measured by the area under the ROC curve (AUROC), with 0.7–0.8, 0.8–0.9, and >0.9 representing acceptable, excellent, and outstanding discrimination, respectively (59–61). Youden's *J* index was used to identify a clinically relevant cut-off score for the PROMIS-A CAT to identify HF patients with moderate to severe anxiety symptoms.

Missing data were not imputed as fewer than 5% of participants were missing any observations used in multivariate adjustment for known-groups comparison. Statistical analyses were performed using Stata version 15.1 and R version 4.3.3.

3 Results

Among the 520 patients enrolled in the main study, 333 completed both PROMIS-A CAT and legacy instruments for validation. Participant characteristics are presented in Table 1. Mean (SD) age was 67 (16) years with 217 (65%) male participants. Of these participants, 87 (26%) had moderate to severe anxiety (GAD-7 score ≥ 10). The most common etiology of HF in this cohort was non-ischemic cardiomyopathy, affecting 218 (66%) participants; those with non-ischemic cardiomyopathy made up a higher proportion of the moderate to severe anxiety cohort (77% vs. 62%, $p = 0.013$). Those with moderate to severe anxiety were younger [mean (SD) age 61 (16) vs. 70 (15) years, $p < 0.001$].

Summary statistics for PROMIS and legacy measures are presented in Table 2. The median (IQR) GAD-7 score was 5 (8), with a higher floor effect (7%) and skewness (0.88) compared to PROMIS-A CAT *T*-scores [median (IQR) 56 (16), floor effect (5%), skewness (−0.19)] (Supplementary Figure S2). The median (IQR) number of PROMIS-A CAT items completed was 4 (1), with a range of 4–12. Seventy-five percent of participants completed ≤ 5 PROMIS-A CAT items.

The mean reliability of PROMIS-A CAT for the total sample was 0.91, with reliability above 0.9 for 87% of participants (see Figure 1 for reliability across all *T*-scores). Cronbach's alpha was 0.9 [95% confidence interval (CI): 0.89–0.92] for the GAD-7, 0.8 (95% CI: 0.82–0.87) for the ESAS-r, and 0.78 (95% CI: 0.75–0.81) for the EQ-5D-5L.

PROMIS-A CAT *T*-scores were strongly correlated with GAD-7 scores ($\rho = 0.78$, 95% CI: 0.73–0.82, $p < 0.001$) and moderately correlated with ESAS-r Anxiety ($\rho = 0.62$, 95% CI: 0.55–0.70, $p < 0.001$) and EQ-5D-Anxiety/Depression ($\rho = 0.67$, 95% CI: 0.60–0.73, $p < 0.001$), as expected (Table 3). PROMIS-A CAT *T*-scores correlated only weakly with constructs unrelated to anxiety, including ESAS-r Appetite ($\rho = 0.18$, 95% CI: 0.07–0.28, $p < 0.001$), EQ-5D-Mobility ($\rho = 0.20$, 95% CI: 0.10–0.31, $p < 0.001$), and KCCQ-Physical limitation ($\rho = -0.27$, 95% CI: −0.38 to −0.16, $p < 0.001$) (Table 3).

TABLE 1 Participant characteristics.

Characteristics	Total (<i>n</i> = 333)	Patients without moderate/severe anxiety (GAD-7 < 10) (<i>n</i> = 246)	Patients with moderate/severe anxiety (GAD-7 ≥ 10) (<i>n</i> = 87)	<i>P</i> -value ^a
Age (years); mean (SD)	67 (16)	70 (15)	61 (16)	<0.001
Sex (male); <i>n</i> (%)	217 (65)	166 (67)	51 (59)	0.14
Marital status; <i>n</i> (%)				0.7
Single	19 (6)	14 (6)	5 (6)	
Married or common law	153 (46)	117 (48)	36 (41)	
Divorced, separated, widowed	14 (4.2)	11 (4.5)	3 (3.4)	
Education; <i>n</i> (%)				0.30
<12 years	39 (13)	32 (15)	7 (9)	
≥12 years	255 (87)	188 (85)	67 (91)	
Smoking (current/former); <i>n</i> (%)	102 (31)	71 (29)	31 (36)	0.2
Serum creatinine (umol/L); mean (SD)	138 (69)	137 (65)	142 (79)	0.9
Blood hemoglobin (g/L); mean (SD)	120 (23)	120 (23)	121 (22)	>0.9
Serum sodium (mmol/L); mean (SD)	137 (5)	137 (4.6)	138 (4.2)	0.3
Serum BNP Level (pg/ml); mean (SD)	1,307 (1,098)	1,265 (1,029)	1,428 (1,278)	0.6
Charlson comorbidity index; <i>n</i> (%)				0.9
<4	288 (86)	213 (87)	75 (86)	
≥4	45 (14)	33 (13)	12 (14)	
LVEF; mean (SD)	37 (17)	37 (17)	35 (17)	0.3
HF type; <i>n</i> (%)				0.6
HFrEF (LVEF <40%)	171 (54)	124 (53)	47 (56)	
HFmEF (LVEF 40%–49%)	47 (15)	33 (14)	14 (17)	
HFpEF (LVEF ≥50%)	101 (32)	78 (33)	23 (27)	
CHF etiology ^b ; <i>n</i> (%)				
CAD/Ischemia	105 (32)	88 (36)	17 (20)	0.006
Valvular	47 (14)	36 (15)	11 (13)	0.7
Hypertension	8 (2.4)	6 (2.4)	2 (2.3)	>0.9
Non-ischemic cardiomyopathy	218 (66)	152 (62)	66 (77)	0.013

GAD-7, General Anxiety Disorder-7; SD, standard deviation; IQR, interquartile range; BNP, B-type natriuretic peptide; LVEF, left ventricular ejection fraction; HFrEF, heart failure with reduced ejection fraction; HFmEF, heart failure with mid-range ejection fraction; HFpEF, heart failure with preserved ejection fraction; HF, heart failure; CHF, congestive heart failure; CAD, coronary artery disease.

^aAlpha threshold for significance after Bonferroni correction for multiple tests = 0.003.

^bCHF etiologies are not mutually exclusive.

TABLE 2 Summary statistics of PROM scores.

PROM	Mean (SD)	Median (IQR)	Cohort score range	Floor effect (%)	Ceiling effect (%)
PROMIS-anxiety <i>T</i> -score v.1.0	55 (11)	56 (16)	33–85	5	1
GAD-7	6 (6)	5 (8)	0–21	17	2
ESAS-r anxiety	3 (3)	1 (5)	0–10	43	2
ESAS-r appetite	3 (3)	1 (5)	0–10	48	3
EQ-5D-5l anxiety/depression	2 (1)	2 (3)	1–5	46	2
PHQ-9	9 (6)	9 (9)	0–27	5	1
KCCQ-12 physical limitation	42 (29)	42 (40)	0–100	11	5
KCCQ-12 summary score	36 (24)	33 (34)	0–98	1	0

PROM, patient-reported outcome measure; SD, standard deviation; IQR, interquartile range; PROMIS, Patient-Reported Outcome Measurement Information System; GAD-7, Generalized Anxiety Disorder-7; ESAS-r, Edmonton Symptom Assessment System Revised; EQ-5D, EuroQol 5-Dimension 5-Level; PHQ-9, Patient-Health Questionnaire-9; KCCQ-12, Kansas City Cardiomyopathy Questionnaire-12.

To further assess construct validity, PROMIS-A CAT and GAD-7 scores were analyzed across groups expected to have different anxiety levels (Table 4). PROMIS-A CAT *T*-scores and GAD-7 scores were both significantly higher among the youngest tertile of participants, and participants with

ESAS-r scores ≥30, KCCQ-12 summary score <25, and PHQ-9 ≥10. Contrary to expectations, PROMIS-A CAT *T*-scores were not significantly different between male and female participants, comorbidity, or HF type. This pattern was similar for GAD-7 scores.

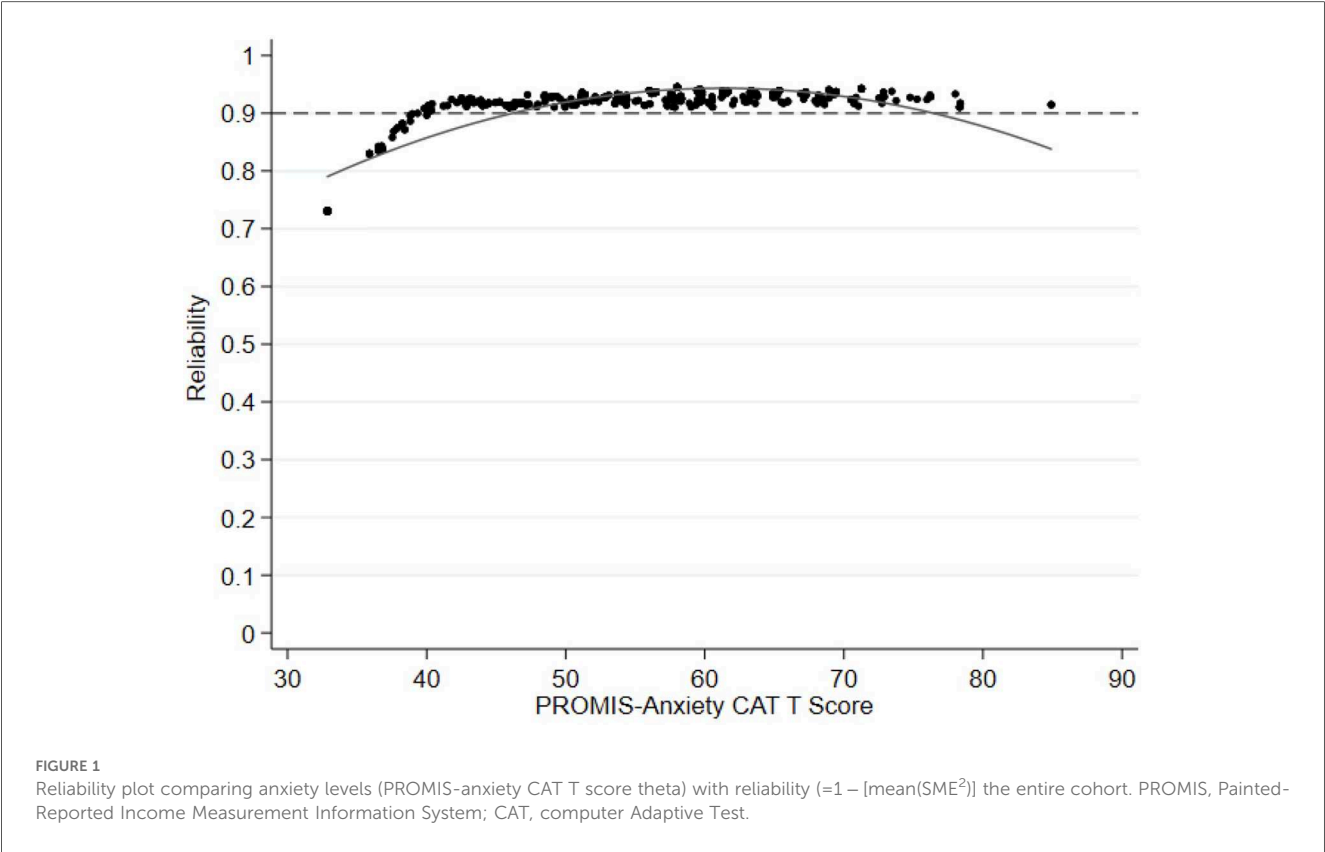


TABLE 3 Convergent and divergent (discriminant) validity of the PROMIS-A CAT T-scores.

PROM	rho	95% Confidence Interval for rho	P-value ^a
GAD-7	0.775	0.730–0.821	<0.001
ESAS-r anxiety	0.624	0.551–0.697	<0.001
EQ-5D-Anxiety/depression	0.668	0.601–0.734	<0.001
ESAS-r appetite	0.176	0.070–0.281	0.001
KCCQ-12 physical limitation	–0.270	(–0.377) to (–0.163)	<0.001
EQ-5D-Mobility	0.204	0.099–0.307	<0.001

For convergent validity, at least moderate correlation ($\rho > 0.6$) was expected with legacy questionnaires measuring the same or similar construct (GAD-7, ESAS-r Anxiety, EQ-5D-Anxiety/Depression). For divergent (discriminant) validity, weak ($\rho < 0.4$) correlation was expected between PROMIS-A CAT T-scores and scores measuring unrelated constructs (ESAS-r Appetite, KCCQ-Physical Limitation, EQ-5D-Mobility). PROMIS-A, Patient-Reported Outcome Measurement Information System Anxiety; CAT, computer adaptive test; GAD-7, Generalized Anxiety Disorder-7; ESAS-r, Edmonton Symptom Assessment System Revised; EQ-5D, EuroQol 5-Dimension; KCCQ-12, Kansas City Cardiomyopathy Questionnaire-12; PROM, patient-reported outcome measure.

^aAlpha threshold for significance after Bonferroni correction for multiple tests = 0.01.

ROC curve analysis showed that PROMIS-A CAT T-scores had excellent discrimination between participants with vs. without moderate to severe anxiety based on GAD-7 ≥ 10 (AUROC: 0.885, 95% CI: 0.846–0.923) (Figure 2). Using Youden’s J Index, PROMIS-A CAT T-scores of ≥ 59 or ≥ 60 were identified as potential thresholds for moderate to severe anxiety [sensitivity = 86% and 85%; specificity = 76% and 78%, respectively (Youden’s $J = 0.63$ for both)] (Supplementary Table S1).

4 Discussion

This study provides evidence supporting the validity and reliability of the PROMIS-A CAT for detecting anxiety symptoms in patients hospitalized for HF. Our results demonstrated very good construct validity, high reliability, and excellent discrimination, which aligns with previous research across other patient populations (25, 62, 63). Given the excellent measurement characteristics, our results establish that PROMIS-A CAT can be considered for use in both clinical practice and research.

We first demonstrated validity of the PROMIS-A CAT when delivered to hospitalized patients with HF by determining its robust convergent validity. The strongest correlation was observed with GAD-7, which is expected since both are multi-item instruments measuring the same construct. The correlation between PROMIS-A (delivered by CAT or SF) and GAD-7 has been consistently strong in other studies (62–64). PROMIS-A CAT demonstrated moderate correlations with the ESAS-r anxiety item and EQ-5D anxiety/depression item. The ESAS-r anxiety assesses anxiety using only one item and has consequently shown only moderate correlation with the GAD-7 (34, 54, 65, 66). Similarly, EQ-5D assesses both anxiety and depression with a single item (38) and exhibits a significant floor effect, with only moderate correlation to other anxiety measures (67). Validity of PROMIS-A CAT T-score was further supported by weak correlation with scores assessing constructs

TABLE 4 Known-group comparisons for PROMIS-A CAT T-scores and GAD-7 scores—adjusted for age, sex, and EF. PROMIS-A CAT T-scores were adjusted in linear regression models using least-squared means function. GAD-7 scores were adjusted in quantile regression models; adjusted medians calculated by predict function. All adjustments performed in R version 4.3.3.

Known-group	PROMIS-A CAT T-score				GAD-7 score	
	Mean (SD)	P-value ^a	Cohen's d	Median (IQR)	P-value ^a	Cliffs' delta ^b
Age (years) ^c n (%)						
19–62 119 (36)	58 (10)	<0.001		7 (11)	<0.001	
63–74 93 (28)	57 (11)			5 (8)		
75–98 121 (36)	51 (10)			2 (6)		
Sex ^d n (%)						
Male 217 (65)	55 (10)	0.675	0.052	5 (7)	0.164	0.094
Female 116 (35)	56 (12)			6 (10)		
Charlson comorbidity index ^e n (%)						
<4 288 (86)	55 (11)	0.727	0.054	5 (7)	0.189	0.126
≥4 45 (14)	56 (10)			6 (6)		
HF type ^f n (%)						
HF _r EF (LVEF < 40%) 171 (54)	55 (11)	0.631		4 (9)	0.105	
HF _m EF (LVEF 40–49%) 47 (15)	56 (9)			6 (9)		
HF _p EF (LVEF ≥ 50%) 101 (32)	55 (10)			5 (6)		
ESAS-r score ^e n (%)						
<30 192 (58)	51 (10)	<0.001	0.929	3 (5)	<0.001	0.465
≥30 140 (42)	60 (9)			8 (9)		
KCCQ-12 summary score ^e n (%)						
<25 109 (37)	60 (10)	<0.001	0.656	8 (10)	<0.001	0.403
≥25 182 (63)	53 (10)			4 (6)		
PHQ-9 Score ^e n (%)						
<10 183 (55)	51 (10)	<0.001	0.945	3 (5)	<0.001	0.597
≥10 149 (45)	60 (9)			9 (9)		

PROMIS-A CAT, Patient-Reported Outcome Measurement Information System anxiety computer adaptive test; GAD-7, Generalized Anxiety Disorder-7; SD, standard deviation; IQR, interquartile range; HF, heart failure; LVEF, left ventricular ejection fraction; HF_rEF, heart failure with reduced ejection fraction; HF_mEF, heart failure with mid-range ejection fraction; HF_pEF, heart failure with preserved ejection fraction; ESAS-r, Edmonton Symptom Assessment System Revised; KCCQ-12, Kansas City Cardiomyopathy Questionnaire-12; PHQ-9, Patient Health Questionnaire-9.

^aAlpha threshold for significance after Bonferroni correction for multiple tests = 0.0,035.
^bEffect size for differences in mean GAD-7 scores were determined using bootstrap resampling of 1,000 replications to account for skewed distribution of GAD-7 scores in sample.
^cMean PROMIS-A CAT T-score/median GAD-7 score adjusted for sex and ejection fraction.
^dMean PROMIS-A CAT T-score/median GAD-7 score adjusted for age and ejection fraction.
^eMean PROMIS-A CAT T-score/median GAD-7 score adjusted for age, sex and ejection fraction.
^fMean PROMIS-A CAT T-score/median GAD-7 score adjusted for age and sex.

unrelated to anxiety [divergent (discriminant) validity]. These correlation coefficients fall within the published range, supporting the robust validity of the PROMIS-A (49).

In known-group comparisons, PROMIS-A CAT T-scores differed between several, but not all, pre-specified sub-groups. Importantly, the distribution pattern was similar for GAD-7, as well, supporting construct validity. As expected, anxiety scores were higher among the youngest tertile of participants (ages: 19–62) vs. the middle (ages: 63–74) and oldest (ages: 75–98) tertiles; those with an ESAS-r score ≥30 (high symptom burden) vs. <30; those with a KCCQ-12 summary score <25 (poor HRQOL) vs. ≥25, and those with a PHQ-9 score ≥10 (moderate/severe depressive symptoms) vs. <10. Contrary to our hypotheses, both PROMIS-A CAT T-scores and GAD-7 scores were similar between female and male participants. However, similar findings have been reported (6, 68). It is possible that the disease severity or other sample characteristics contributed to this result. Similarly, no difference was detected between groups formed by comorbidity or HF type. Of note, those with non-ischemic cardiomyopathy made up a larger proportion of the moderate-

severe anxiety cohort based on the GAD-7 ≥10. This can be attributed to younger average age among patients with this etiology, compared to coronary artery disease or ischemic etiologies that are more prevalent in older populations (69).

Our results showed excellent reliability of the PROMIS-A CAT, with individual-level reliability demonstrated by minimal standard error of measurements across the spectrum of PROMIS-A T-scores. This standard of high reliability is a built-in strength of PROMIS CAT, as standard stopping rules require a low standard error (theta <0.3) to stop additional testing (24, 25). Individual-level reliability decreased for participants with the least anxiety (lowest T-scores). This is not of clinical relevance, particularly if the tool is used for screening, as this score range falls well below the cut-off for potentially clinically significant symptom severity.

The PROMIS-A CAT exhibited excellent coverage across the range of anxiety severity, demonstrating no significant ceiling effect and a smaller floor effect compared to legacy measures. The floor effect for CAT administration in this study was lower than that observed when the item bank was administered as a

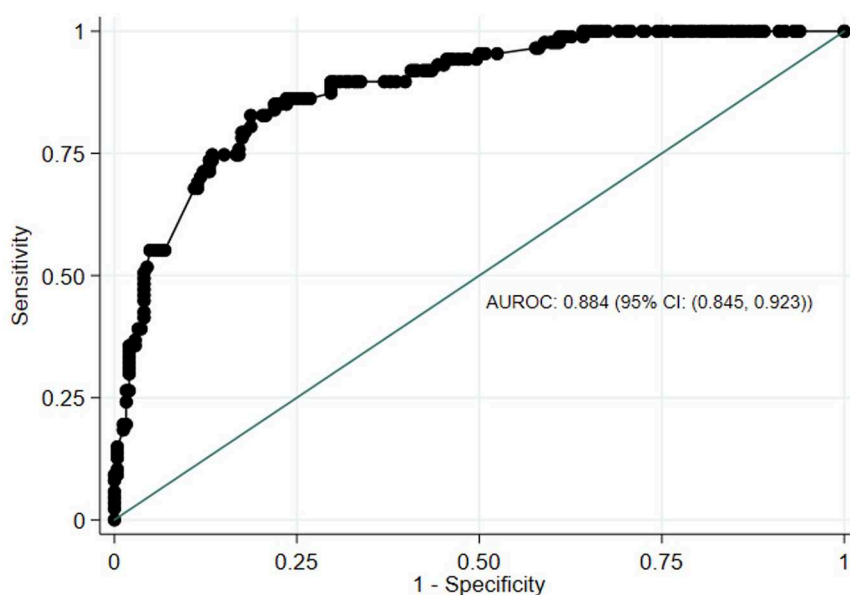


FIGURE 2

Receiver-operating characteristics (ROC) curve of PROMIS-anxiety CAT *T* scores against GAD-7 scores. AUROC, area under receiving-operating characteristics. PROMIS, Painted-Reported Income Measurement Information System; CAT, computer Adaptive Test; GAD-7, Generalized Anxiety Disorder-7; CI, Confidence Interval.

SF in other patient populations (70–72), demonstrating a more tailored assessment with CAT (73).

PROMIS-A CAT demonstrated excellent discrimination between participants with vs. without moderate/severe anxiety. Our threshold analysis demonstrated near-identical specificity and sensitivity when using a cut-off score of 59 or 60, consistent with studies in other patient populations (64, 74). We recommend a cut-off *T*-score of ≥ 60 to identify patients with HF who may benefit from further assessment for potential moderate/severe anxiety. This threshold is one standard deviation above the reference value for the United States general population, which is congruent with the suggested distribution-based cut-off for moderate symptom burden across many PROMIS domains (75, 76).

Most participants in this study completed only 4–5 items with the PROMIS-A CAT to obtain a reliable score, compared to the seven required items with the GAD-7. This number of items is similar to findings in other patient populations completing PROMIS item banks via CAT (23, 77). On average, participants with no anxiety were required to answer more questions than those with higher *T*-scores. This is because these participants usually answered “never” to multiple items, which conveys insufficient information for the CAT to reach a low enough SEM to fulfill the stopping rule (78). This may elicit frustration; however, this can be addressed by modifying CAT stopping rules to limit the maximum number of items administered to 6 or 8, without losing precision. An even more efficient solution is a recent, optional modification, referred to as the “screen-to-CAT” method. If a participant selects “never” for the first item, no further question is asked (24, 79). Since the PROMIS items

are calibrated based on the item response theory, even a single answer will yield a sufficiently reliable *T*-score.

Prior research has identified 7 core PROMIS domains (anxiety, depression, fatigue, pain interference, physical function, sleep disturbance, social functions), which greatly contribute to HRQOL (80, 81) and are relevant across many chronic conditions (82). Since PROMIS tools are not disease specific, they can be used to measure and compare these domains across many conditions (83). Based on average response times, it is possible to assess these 7 domains in 5–10 min, making PROMIS tools ideal candidates for symptom screening (25). Additionally, domain-specific *T*-scores can be combined to generate mental and physical health summary scores to characterize overall HRQOL on population level (84). Furthermore, PROMIS domain *T*-scores can be combined into a preference-based health utility score (PROMIS Preference score; PROPr), which provides an overall measure of patient quality of life and serves as a helpful metric for health economy analyses (85, 86).

PROMs and PROMIS tools are increasingly considered for clinical use, primarily for symptom assessment and monitoring. Major electronic medical record platforms now include PROM modules and PROMIS tools, allowing clinicians to efficiently track patient data and respond to changes in health status. This integration streamlines workflows, consolidates data, and enables patients to access their results through patient portals, empowering them to actively track and co-manage their own health parameters (87). A recent implementation trial showed feasibility when integrating PROMIS symptom scores into the electronic medical records of ambulatory oncology patients (88).

The results of this study should be interpreted in context of some limitations. We recruited a convenience sample of hospitalized patients with HF, which may not be representative of all hospitalized HF patients. Patients who felt more comfortable with electronics were likely overrepresented in this sample given the PROMIS-A CAT delivery method. Though, this should not impact the conclusions regarding validity and reliability of the PROMIS-A CAT *T*-scores. Nevertheless, computer literacy is an important consideration before implementation of CAT (25). Additionally, this analysis was cross-sectional and did not assess responsiveness of the PROMIS-A CAT; longitudinal validation is a focus for future work. Non-English speakers were excluded from recruitment; future studies should validate PROMIS-A CAT using translated item banks. Finally, while our study assessed the measurement characteristics of PROMIS-A CAT, the most efficient strategies for implementing the tool into clinical care remain unknown. Future implementation studies are needed to evaluate the feasibility of routine PROMIS-A CAT administration to screen and monitor for anxiety among patients with HF. Important considerations for these studies include assessing individual- and system-level barriers and facilitators toward adoption in the clinical context, understanding patient's and clinician's perceived utility of the instrument, and evaluating the impact of implementation on outcomes such as symptom burden, HRQOL, and healthcare use. It is important to note that in such studies, in addition to administering PROMIS-A CAT as a screening tool, it is essential for success that appropriate evidence-based symptom management pathways are available. Since the PROMIS tools are not diagnostic tools, individuals with potentially significant anxiety symptoms should be assessed by a qualified professional. If needed, resources for appropriate interventions, which may include, but are not limited to, self-care strategies, social support, psychotherapy, and medications, will need to be available (89).

5 Conclusion

In conclusion, we provide evidence supporting the reliability and validity of the PROMIS CAT anxiety item bank for hospitalized patients with HF. We recommend a cut-off score of ≥ 60 to identify patients with moderate/severe anxiety symptoms who may benefit from further clinical assessment. The PROMIS-A CAT demonstrates high measurement precision with fewer items than traditional legacy measures, making it an efficient tool for screening.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University Health Network Research Ethics Board. The studies were

conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

NM: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. SC: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. NT: Conceptualization, Writing – original draft, Writing – review & editing. DL: Conceptualization, Data curation, Writing – original draft, Writing – review & editing. AB: Conceptualization, Supervision, Writing – original draft, Writing – review & editing. NE: Data curation, Formal analysis, Methodology, Supervision, Writing – review & editing. HR: Writing – review & editing. JS: Writing – review & editing. DT: Data curation, Writing – original draft, Writing – review & editing. JF: Formal analysis, Methodology, Writing – review & editing. JL: Writing – original draft, Writing – review & editing. IM: Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The “Predicting Readmission Outcomes using Biostatistical Evaluation and Machine Learning (PROBE ML)” study is funded by the Ted Rogers Centre for Heart Research and the Heart and Stroke Foundation of Canada.

Conflict of interest

IM received non-restricted education grants from Astellas Canada and from Paladin Labs Canada.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Heart Failure in Canada: Complex, Incurable, and on the Rise. Toronto, ON: Heart and Stroke Foundation of Canada (2022).
2. Schwinger RHG. Pathophysiology of heart failure. *Cardiovasc Diagn Ther.* (2021) 11(1):263–76. doi: 10.21037/cdt-20-302
3. Easton K, Coventry P, Lovell K, Carter L-A, Deaton C. Prevalence and measurement of anxiety in samples of patients with heart failure: meta-analysis. *J Cardiovasc Nurs.* (2016) 31(4):367–79. doi: 10.1097/JCN.0000000000000265
4. Watson RD, Gibbs CR, Lip GY. ABC of heart failure. Clinical features and complications. *Br Med J.* (2000) 320(7229):236–9. doi: 10.1136/bmj.320.7229.236
5. Friedmann E, Thomas SA, Liu F, Morton PG, Chapa D, Gottlieb SS. Relationship of depression, anxiety, and social isolation to chronic heart failure outpatient mortality. *Am Heart J.* (2006) 152(5):940.e1–8. doi: 10.1016/j.ahj.2006.05.009
6. Costa FMD, Martins SPV, Moreira ECTD, Cardoso JCMS, Fernandes LPNS. Anxiety in heart failure patients and its association with socio-demographic and clinical characteristics: a cross-sectional study. *Porto Biomed J.* (2022) 7(4):e177. doi: 10.1097/j.pbj.0000000000000177
7. Tsabedze N, Kinsey J-LH, Mpanya D, Mogashoa V, Klug E, Manga P. The prevalence of depression, stress and anxiety symptoms in patients with chronic heart failure. *Int J Ment Health Syst.* (2021) 15(1):44. doi: 10.1186/s13033-021-00467-x
8. Bean MK, Gibson D, Flattery M, Duncan A, Hess M. Psychosocial factors, quality of life, and psychological distress: ethnic differences in patients with heart failure. *Prog Cardiovasc Nurs.* (2009) 24(4):131–40. doi: 10.1111/j.1751-7117.2009.00051.x
9. Celano CM, Villegas AC, Albanese AM, Gaggin HK, Huffman JC. Depression and anxiety in heart failure: a review. *Harv Rev Psychiatry.* (2018) 26(4):175–84. doi: 10.1097/HRP.0000000000000162
10. Kelkar AA, Spertus J, Pang P, Pierson RF, Cody RJ, Pina IL, et al. Utility of patient-reported outcome instruments in heart failure. *JACC Heart Fail.* (2016) 4(3):165–75. doi: 10.1016/j.jchf.2015.10.015
11. Zannad F, Alikhaani J, Alikhaani S, Butler J, Gordon J, Jensen K, et al. Patient-reported outcome measures and patient engagement in heart failure clinical trials: multi-stakeholder perspectives. *Eur J Heart Fail.* (2023) 25(4):478–87. doi: 10.1002/ehf.2828
12. Mackintosh A, Gibbons E, Fitzpatrick R. *A Structured Review of Patient-Reported Outcome Measures (PROMs) for Heart Failure.* Oxford: University of Oxford (2009).
13. Schougaard LM, Larsen LP, Jessen A, Sidenius P, Dorflinger L, de Thurah A, et al. Ambuflex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. *Qual Life Res.* (2016) 25(3):525–34. doi: 10.1007/s11366-015-1207-0
14. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol.* (2016) 34(6):557–65. doi: 10.1200/JCO.2015.63.0830
15. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* (2006) 166(10):1092–7. doi: 10.1001/archinte.166.10.1092
16. Chamberlain AM. Editorial commentary: legacy patient-reported outcome measures remain important today despite responder burden, but with further refinement, patient-reported outcomes measurement information system could replace legacy instruments in the future. *Arthroscopy.* (2023) 39(3):853–5. doi: 10.1016/j.arthro.2022.08.032
17. Lynch CP, Cha EDK, Jadcak CN, Mohan S, Geoghegan CE, Singh K. What can legacy patient-reported outcome measures tell US about participation bias in patient-reported outcomes measurement information system scores among lumbar spine patients? *Neurospine.* (2022) 19(2):307–14. doi: 10.14245/ns.2040706.353
18. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med Care.* (2007) 45(5 Suppl 1):S3–11. doi: 10.1097/01.mlr.0000258615.42478.55
19. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcvm.2025.1605130/full#supplementary-material>

- reported outcomes measures. *Clin Ther.* (2014) 36(5):648–62. doi: 10.1016/j.clinthera.2014.04.006
20. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol.* (2005) 23(5 Suppl 39):S53–7.
21. Health Measures. Intro to PROMIS® (2024). Available online at: <https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis> (Accessed May 16, 2024).
22. Kimura T. The impacts of computer adaptive testing from a variety of perspectives. *J Educ Eval Health Prof.* (2017) 14:12. doi: 10.3352/jeehp.2017.14.12
23. Dano S, Hussain J, Edwards N, Sun YI, Li M, Howell D, et al. Assessing fatigue in patients receiving kidney replacement therapy using PROMIS computer adaptive testing. *Am J Kidney Dis.* (2023) 82(1):33–42.e1. doi: 10.1053/j.ajkd.2022.12.018
24. Health Measures. Stopping rules. PROMIS (2024). Available online at: <https://www.healthmeasures.net/resource-center/measurement-science/computer-adaptive-tests-cats/stopping-rules> (Accessed August 1, 2024).
25. van der Willik EM, van Breda F, van Jaarsveld BC, van de Putte M, Jetten IW, Dekker FW, et al. Validity and reliability of the patient-reported outcomes measurement information system (PROMIS(R)) using computerized adaptive testing in patients with advanced chronic kidney disease. *Nephrol Dial Transplant.* (2023) 38(5):1158–69. doi: 10.1093/ndt/gfac231
26. Lapin B, Davin S, Stilphen M, Benzel E, Katzan IL. Validation of PROMIS CATs and PROMIS global health in an interdisciplinary pain program for patients with chronic low back pain. *Spine (Phila Pa 1976).* (2020) 45(4):E227–35. doi: 10.1097/BRS.0000000000000323
27. Rafiq RB, Yount S, Jerousek S, Roth EJ, Cella D, Albert MV, et al. Feasibility of PROMIS using computerized adaptive testing during inpatient rehabilitation. *J Patient Rep Outcomes.* (2023) 7(1):44. doi: 10.1186/s41687-023-00567-x
28. Cao Z, Jia Y, Zhu B. BNP and NT-proBNP as diagnostic biomarkers for cardiac dysfunction in both clinical and forensic medicine. *Int J Mol Sci.* (2019) 20(8):1820. doi: 10.3390/ijms20081820
29. Allen LA, Hernandez AF, Peterson ED, Curtis LH, Dai D, Masoudi FA, et al. Discharge to a skilled nursing facility and subsequent clinical outcomes among older patients hospitalized for heart failure. *Circ Heart Fail.* (2011) 4(3):293–300. doi: 10.1161/CIRCHEARTFAILURE.110.959171
30. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. *J Clin Epidemiol.* (2016) 73:119–27. doi: 10.1016/j.jclinepi.2015.08.036
31. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. 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(11):1179–94. doi: 10.1016/j.jclinepi.2010.04.011
32. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res.* (1997) 6(6):595–600. doi: 10.1023/A:1018420418455
33. Ruiz MA, Zamorano E, Garcia-Campayo J, Pardo A, Freire O, Rejas J. Validity of the GAD-7 scale as an outcome measure of disability in patients with generalized anxiety disorders in primary care. *J Affect Disord.* (2011) 128(3):277–86. doi: 10.1016/j.jad.2010.07.010
34. Hui D, Bruera E. The Edmonton symptom assessment system 25 years later: past, present, and future developments. *J Pain Symptom Manage.* (2017) 53(3):630–43. doi: 10.1016/j.jpainsymman.2016.10.370
35. Watanabe SM, Nikolaichuk C, Beaumont C, Johnson L, Myers J, Strasser F. A multicenter study comparing two numerical versions of the Edmonton symptom assessment system in palliative care patients. *J Pain Symptom Manage.* (2011) 41(2):456–68. doi: 10.1016/j.jpainsymman.2010.04.020
36. Opasich C, Gualco A, De Feo S, Barbieri M, Cioffi G, Giardini A, et al. Physical and emotional symptom burden of patients with end-stage heart failure: what to measure, how and why. *J Cardiovasc Med.* (2008) 9(11):1104–8. doi: 10.2459/JCM.0b013e32830c1b45
37. Kularatna S, Senanayake S, Chen G, Parsonage W. Mapping the Minnesota living with heart failure questionnaire (MLHFQ) to EQ-5D-5l in patients with

- heart failure. *Health Qual Life Outcomes*. (2020) 18(1):115. doi: 10.1186/s12955-020-01368-2
38. Thomas M, Jones PG, Cohen DJ, Suzanne AV, Magnuson EA, Wang K, et al. Predicting the EQ-5D utilities from the Kansas city cardiomyopathy questionnaire in patients with heart failure. *Eur Heart J Qual Care Clin Outcomes*. (2021) 7(4):388–96. doi: 10.1093/ehjqcco/qcab014
39. Devlin N, Parkin D, Janssen B. *Methods for Analysing and Reporting EQ-5D Data*. Cham: Springer (2020).
40. Gao Z, Wang P, Hong J, Yan Y, Tong T, Wu B, et al. Health-related quality of life among Chinese patients with crohn's disease: a cross-sectional survey using the EQ-5D-5L. *Health Qual Life Outcomes*. (2022) 20(1):62. doi: 10.1186/s12955-022-01969-z
41. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. (2001) 16(9):606–13. doi: 10.1046/j.1525-1497.2001.016009606.x
42. Stubblefield WB, Jenkins CA, Liu D, Storrow AB, Spertus JA, Pang PS, et al. Improvement in Kansas city cardiomyopathy questionnaire scores after a self-care intervention in patients with acute heart failure discharged from the emergency department. *Circ Cardiovasc Qual Outcomes*. (2021) 14(10):e007956. doi: 10.1161/CIRCOUTCOMES.121.007956
43. Haynes W. Bonferroni correction. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, editors. *Encyclopedia of Systems Biology*. New York: Springer (2013). p. 154.
44. Kovchegov Y. A new life of Pearson's skewness. *J Theor Probab*. (2022) 35:2896–915. doi: 10.1007/s10959-021-01149-7
45. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care*. (2000) 38(9 Suppl):I128–42. doi: 10.1097/00005650-200009002-00007
46. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. (2011) 2:53–5. doi: 10.5116/ijme.4dfb.8dfd
47. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. Boston, MA: Houghton Mifflin (2003).
48. Konstam V, Moser DK, De Jong MJ. Depression and anxiety in heart failure. *J Card Fail*. (2005) 11(6):455–63. doi: 10.1016/j.cardfail.2005.03.006
49. Uszko-Lencer N, Mesquita R, Janssen E, Werter C, Brunner-La Rocca H-P, Pitta F, et al. Reliability, construct validity and determinants of 6-minute walk test performance in patients with chronic heart failure. *Int J Cardiol*. (2017) 240:285–90. doi: 10.1016/j.ijcard.2017.02.109
50. Farhane-Medina NZ, Luque B, Tabernero C, Castillo-Mayén R. Factors associated with gender and sex differences in anxiety prevalence and comorbidity: a systematic review. *Sci Prog*. (2022) 105(4):368504221135469. doi: 10.1177/00368504221135469
51. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch Gen Psychiatry*. (2005) 62(6):593–602. doi: 10.1001/archpsyc.62.6.593
52. AbuRuz ME. Anxiety and depression predicted quality of life among patients with heart failure. *J Multidiscip Healthc*. (2018) 11:367–73. doi: 10.2147/JMDH.S170327
53. Dai D, Coetzer H, Zion SR, Malecki MJ. Multimorbidity and its associations with anxiety and depression among newly diagnosed patients with breast cancer: a retrospective observational cohort study in a US commercially insured and medicare advantage population. *Cancer Control*. (2022) 29:10732748221140691. doi: 10.1177/10732748221140691
54. Tang E, Dano S, Edwards N, Macanovic S, Ford H, Bartlett S, et al. Screening for symptoms of anxiety and depression in patients treated with renal replacement therapy: utility of the Edmonton symptom assessment system-revised. *Qual Life Res*. (2022) 31(2):597–605. doi: 10.1007/s11136-021-02910-5
55. Spertus JA, Jones PG, Sandhu AT, Arnold SV. Interpreting the Kansas city cardiomyopathy questionnaire in clinical trials and clinical care: JACC state-of-the-art review. *J Am Coll Cardiol*. (2020) 76(20):2379–90. doi: 10.1016/j.jacc.2020.09.542
56. Ruscio J. A probability-based measure of effect size: robustness to base rates and other factors. *Psychol Methods*. (2008) 13(1):19–30. doi: 10.1037/1082-989X.13.1.19
57. Meissel K, Yao E. Using cliff's Delta as a non-parametric effect size measure: an accessible web app and R tutorial. *Pract Assess*. (2024) 29:2. doi: 10.7275/pare.1977
58. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item generalized anxiety disorder questionnaire (GAD-7). *J Affect Disord*. (2020) 265:395–401. doi: 10.1016/j.jad.2020.01.032
59. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *Jama*. (2017) 318(14):1377–84. doi: 10.1001/jama.2017.12126
60. Mandrekari JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. (2010) 5(9):1315–6. doi: 10.1097/JTO.0b013e3181ec173d
61. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J Anesthesiol*. (2022) 75(1):25–36. doi: 10.4097/kja.21209
62. Clover K, Lambert SD, Oldmeadow C, Britton B, Mitchell AJ, Carter G, et al. Convergent and criterion validity of PROMIS anxiety measures relative to six legacy measures and a structured diagnostic interview for anxiety in cancer patients. *J Patient Rep Outcomes*. (2022) 6(1):80. doi: 10.1186/s41687-022-00477-4
63. Saqib M, Yang M, Jamal F, Aghamohammadi S, Ahmadzadeh G, Hamid M, et al. *Validation of PROMIS Anxiety Computer Adaptive Test in Solid Organ Transplant Recipients*. Mt. Laurel, NJ: American Transplant Congress (2021).
64. Purvis TE, Neuman BJ, Riley LH, Skolasky RL. Comparison of PROMIS anxiety and depression, PHQ-8, and GAD-7 to screen for anxiety and depression among patients presenting for spine surgery. *J Neurosurg Spine*. (2019) 30(4):524–31. doi: 10.3171/2018.9.SPINE18521
65. Dano S, Pokarowski M, Liao B, Tang E, Ekundayo O, Li V, et al. Evaluating symptom burden in kidney transplant recipients: validation of the revised Edmonton symptom assessment system for kidney transplant recipients—a single-center, cross-sectional study. *Transpl Int*. (2020) 33(4):423–36. doi: 10.1111/tri.13572
66. Richardson LA, Jones GW. A review of the reliability and validity of the Edmonton symptom assessment system. *Curr Oncol*. (2009) 16(1):55. doi: 10.3747/co.v16i1.261
67. Franklin M, Enrique A, Palacios J, Richards D. Psychometric assessment of EQ-5D-5L and ReQoL measures in patients with anxiety and depression: construct validity and responsiveness. *Qual Life Res*. (2021) 30(9):2633–47. doi: 10.1007/s11136-021-02833-1
68. Muller-Tasch T, Müller-Tasch T, Löwe B, Lossnitzer N, Frankenstein L, Täger T, et al. Anxiety and self-care behaviour in patients with chronic systolic heart failure: a multivariate model. *Eur J Cardiovasc Nurs*. (2018) 17(2):170–7. doi: 10.1177/1474515117722255
69. Tsiouris A, Borgi J, Karam J, Nemeh HW, Paone G, Brewer RJ, et al. Ischemic versus nonischemic dilated cardiomyopathy: the implications of heart failure etiology on left ventricular assist device outcomes. *ASAIO J*. (2013) 59(2):130–5. doi: 10.1097/MAT.0b013e31828579af
70. Beleckas CM, Prather H, Guattery J, Wright M, Kelly M, Calfee RP. Anxiety in the orthopedic patient: using PROMIS to assess mental health. *Qual Life Res*. (2018) 27(9):2275–82. doi: 10.1007/s11136-018-1867-7
71. Driban JB, Morgan N, Price LL, Cook KF, Wang C. Patient-reported outcomes measurement information system (PROMIS) instruments among individuals with symptomatic knee osteoarthritis: a cross-sectional study of floor/ceiling effects and construct validity. *BMC Musculoskelet Disord*. (2015) 16:253. doi: 10.1186/s12891-015-0715-y
72. Liu W, Dindo L, Hadlandsmyth K, Unick GJ, Bridget Zimmerman M, Marie BS, et al. Item response theory analysis: PROMIS(R) anxiety form and generalized anxiety disorder scale. *West J Nurs Res*. (2022) 44(8):765–72. doi: 10.1177/01939459211015985
73. Segawa E, Schalet B, Cella D. A comparison of computer adaptive tests (CATs) and short forms in terms of accuracy and number of items administered using PROMIS profile. *Qual Life Res*. (2020) 29(1):213–21. doi: 10.1007/s11136-019-02312-8
74. Cheng AL, Downs DL, Brady BK, Hong BA, Park P, Prather H, et al. Interpretation of PROMIS depression and anxiety measures compared with DSM-5 diagnostic criteria in musculoskeletal patients. *JBJS Open Access*. (2023) 8(1):e22.00110. doi: 10.2106/JBJS.OA.22.00110
75. van Muilekom MM, Luijten MAJ, van Oers HA, Terwee CB, van Litsenburg RRL, Roorda LD, et al. From statistics to clinics: the visual feedback of PROMIS(R) CATs. *J Patient Rep Outcomes*. (2021) 5(1):55. doi: 10.1186/s41687-021-00324-y
76. Health Measures. PROMIS® score cut points (2023). Available online at: <https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points> (Accessed August 1, 2024).
77. Hussain J, Chawla G, Rafiqzad H, Huang S, Bartlett SJ, Howell D, et al. Validation of the PROMIS sleep disturbance item bank computer adaptive test (CAT) in patients on renal replacement therapy. *Sleep Med*. (2022) 90:36–43. doi: 10.1016/j.sleep.2022.01.001
78. You DS, Cook KF, Domingue BW, Ziadni MS, Hah JM, Darnall BD, et al. Customizing CAT administration of the PROMIS misuse of prescription pain medication item bank for patients with chronic pain. *Pain Med*. (2021) 22(7):1669–75. doi: 10.1093/pm/pnab159
79. Health Measures. Anxiety measure differences. PROMIS (2023). Available online at: https://www.healthmeasures.net/images/PROMIS/Differences_Between_PROMIS_Measures/PROMIS_Anxiety_Measure_Differences_08Sept2023.pdf (Accessed August 1, 2024).
80. Samartzis L, Dimopoulos S, Tziogourou M, Nanas S. Effect of psychosocial interventions on quality of life in patients with chronic heart failure: a meta-analysis of randomized controlled trials. *J Card Fail*. (2013) 19(2):125–34. doi: 10.1016/j.cardfail.2012.12.004

81. Quevedo HC, Deravil D, Seo DM, Hebert KA. The meaningful use of the review of symptoms in heart failure patients. *Congest Heart Fail.* (2011) 17(1):31–7. doi: 10.1111/j.1751-7133.2010.00203.x
82. Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, et al. PROMIS(R) adult health profiles: efficient short-form measures of seven health domains. *Value Health.* (2019) 22(5):537–44. doi: 10.1016/j.jval.2019.02.004
83. Flynn KE, Dew MA, Lin L, Fawzy M, Graham FL, Hahn EA, et al. Reliability and construct validity of PROMIS(R) measures for patients with heart failure who undergo heart transplant. *Qual Life Res.* (2015) 24(11):2591–9. doi: 10.1007/s11136-015-1010-y
84. Huang W, Rose AJ, Bayliss E, Baseman L, Butcher E, Garcia R-E, et al. Adapting summary scores for the PROMIS-29 v2.0 for use among older adults with multiple chronic conditions. *Qual Life Res.* (2019) 28(1):199–210. doi: 10.1007/s11136-018-1988-z
85. Klapproth CP, Sidey-Gibbons CJ, Valderas JM, Rose M, Fischer F. Comparison of the PROMIS preference score (PROPr) and EQ-5D-5l index value in general population samples in the United Kingdom, France, and Germany. *Value Health.* (2022) 25(5):824–34. doi: 10.1016/j.jval.2021.10.012
86. Dewitt B, Jalal H, Hanmer J. Computing PROPr utility scores for PROMIS(R) profile instruments. *Value Health.* (2020) 23(3):370–8. doi: 10.1016/j.jval.2019.09.2752
87. Nolla K, Rasmussen LV, Rothrock NE, Butt Z, Bass M, Davis K, et al. Seamless integration of computer-adaptive patient reported outcomes into an electronic health record. *Appl Clin Inform.* (2024) 15(1):145–54. doi: 10.1055/a-2235-9557
88. Penedo FJ, Medina HN, Moreno PI, Sookdeo V, Natori A, Boland C, et al. Implementation and feasibility of an electronic health record-integrated patient-reported outcomes symptom and needs monitoring pilot in ambulatory oncology. *JCO Oncol Pract.* (2022) 18(7):e1100–13. doi: 10.1200/OP.21.00706
89. Rashid S, Qureshi AG, Noor TA, Yaseen K, Sheikh MAA, Malik M, et al. Anxiety and depression in heart failure: an updated review. *Curr Probl Cardiol.* (2023) 48(11):101987. doi: 10.1016/j.cpcardiol.2023.101987