2022

Improving Interpretation of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Scale for Specific Tasks in Community-Dwelling Older Adults

Jeff Houck

Ryan Jacobson

Michael Bass

Chris Dasilva

Judith F. Baumhauer

Improving Interpretation of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Scale for Specific Tasks in Community-Dwelling Older Adults

Jeff Houck, PT, PhD[1]; Ryan Jacobson, PT, DPT[1]; Michael Bass, MS[2]; Chris Dasilva, BS[3]; Judith F. Baumhauer, MD, MPH[3]

ABSTRACT

Background and Purpose: New generic patient-reported outcomes like the Patient-Reported Outcomes Measurement Information System (PROMIS) are available to physical therapists to assess physical function. However, the interpretation of the PROMIS Physical Function (PF) T-score is abstract because it references the United States average and not specific tasks. The purposes of this study were to (1) determine convergent validity of the PROMIS PF scale with physical performance tests; (2) compare predicted performance test values to normative data; and (3) identify sets of PROMIS PF items similar to performance tests that also scale in increasing difficulty and align with normative data.
Methods: Community-dwelling older adults (n = 45; age = 77.1 ± 4.6 years) were recruited for this cross-sectional analysis of PROMIS PF and physical performance tests. The modified Physical Performance Test (mPPT), a multicomponent test of mostly timed items, was completed during the same session as the PROMIS PF scale. Regression analysis examined the relationship of mPPT total and component scores (walking velocity, stair ascent, and 5 times sit to stand) with the PROMIS PF scale T-scores. Normative data were compared with regression-predicted mPPT timed performance across PROMIS PF T-scores. The PROMIS PF items most similar to walking, stair ascent, or sit to stand were identified and then PROMIS PF model parameter–calibrated T-scores for these items were compared alongside normative data.
Results and Discussion: There were statistically significant correlations (r = 0.32-0.64) between PROMIS PF T-score and mPPT total and component scores. Regression-predicted times for walking, stair ascent, and sit-to-stand tasks (based on T-scores) aligned with published normative values for older adults. Selected PF items for stair ascent and walking scaled well to discriminate increasing difficulty; however, sit-to-stand items discriminated only lower levels of functioning.
Conclusions: The PROMIS PF T-scores showed convergent validity with physical performance and aligned with published normative data. While the findings are not predictive of individual performance, they improve clinical interpretation by estimating a range of expected performance for walking, stair ascent, and sit to stand. These findings support application of T-scores in physical therapy testing, goal setting, and wellness plans of care for community-dwelling older adults.
Key Words: patient-reported outcomes, physical function, PROMIS

INTRODUCTION

New generic patient-reported outcome (PRO) scales such as the Patient-Reported Outcomes Measurement Information System (PROMIS) are demonstrating validity equivalent to or better than other generic and disease-specific PRO scales.[1-6] Generic PRO assessment decreases the burden on patients and providers by using computer-adaptive testing (CAT) and minimizing the need for multiple disease-specific scales,[1,7] accelerating adoption of generic PRO instruments.[8]

The generic PROMIS Physical Function (PF) CAT offers many potential advantages to physical therapists for the assessment of patient physical abilities.[9] The 124 PF items in version 1.0 comprise various functional activities, each identified through patient and expert focus groups and then individually calibrated using item response theory.[9,10] The PROMIS PF CAT asks patients to rate their level of

[1]Doctor of Physical Therapy Program, George Fox University, Newberg, Oregon.

[2]Department of Medical Social Sciences, Northwestern University, Chicago, Illinois.

[3]School of Medicine and Dentistry, Department of Orthopaedics, University of Rochester, Rochester, New York.

The authors declare no conflicts of interest.

Address correspondence to: Ryan Jacobson, PT, DPT, Doctor of Physical Therapy Program George Fox University, 501 Villa Rd, Newberg, OR 97132 (rjacobson@georgefox.edu).

The Decision Editor was Robert Wellmon.

difficulty for select items and then chooses appropriately difficult items (adapted to each previous response), avoiding floor and ceiling effects.[1,2,5,6,11] Each PF item offers a 5-point Likert scale with responses such as "no difficulty," "little difficulty," "some difficulty," "much difficulty," and "unable to do." The CAT algorithms determine the patient's PF score in relatively few responses (4-10) and very little time (<1 minute).[5,9,10] This limits patient burden and yet still allows for prediction of difficulty rating on *all 124 items* (due to individual item calibration).[9,10] These advantages mean that the PROMIS PF CAT can be used for patients with a wide range of functional abilities.[12]

Despite such advantages, the PROMIS PF CAT scale produces a T-score for which interpretation is not intuitive. The T-score is normed such that 50 represents the average physical function for the general US population and 10 points is equivalent to 1 standard deviation.[12] A higher or lower T-score indicates better or worse PF, respectively. T-scores are reported along with percentile ranks relative to the general US population, age, and gender. Community-dwelling older adults show T-scores equivalent to the US average,[13] while lower T-scores are reported for clinical populations and cohorts in assisted living.[14] The larger clinical issue is that the T-score does not reveal which of the PROMIS PF items the patient finds difficult. For example, a provider knows that a patient with a T-score of 40 is 1 standard deviation below the US average but not what tasks the patient finds difficult. This contrasts with many disease-specific scales where all individual items are answered and easily referenced.

An improvement in the utility of the PROMIS PF CAT in physical therapy practice would be the ability to link the T-score value with tasks a patient finds difficult.[9] One way to do this would be to demonstrate convergent validity of the PROMIS PF CAT with physical performance measures. A recent study of younger individuals (aged 18-50 years) with hip injuries demonstrated convergence between the PROMIS PF CAT and physical performance measures ($r > 0.70$).[15] However, studies of community-dwelling older adults[16-18] and individuals after surgery[19-23] showed lower convergence (r values between 0.22 and 0.67) when comparing various PRO instruments and physical performance. The level of convergence of current PRO data with physical performance ($r < 0.67$) at best will provide estimates of physical performance that are not precise. However, the errors associated with these estimates may be sufficiently narrow to discriminate a meaningful range of expected physical performance clinically (eg, predicting gait speed above 1.0 m/s).

A second way to increase utility would be to determine alignment of T-scores with normative data for specific performance tests. Three specific timed performance tasks—walking velocity, stair ascent, and 5 times sit to stand (5× STS)—have published norms for community-dwelling older adults.[24-26] Because PROMIS PF CAT scores are based on a large pool of participants (ie, 2010 US census),[2,11] the T-scores would be expected to predict performance (eg, time ascending stairs) that compares with norms for community-dwelling older adults.

Finally, identifying sets of increasingly difficult individual items from the PROMIS PF CAT item bank that are similar to each task (ie, walking, stair ascent, or sit to stand) may also improve utility in using and interpreting T-scores. Previous work has sorted PROMIS PF items into International Classification of Functioning, Disability and Health (ICF) Activities and Participation codes (eg, d450: walking), defining sets of PF items that are similar to specific tasks.[27,28] Ideally, sets of PF items that are matched by ICF coding to a given physical task (eg, walking) exist that also scale in increasing difficulty, such that an average community-dwelling older adult (T-score = 50) would likely self-report "no difficulty" for the easiest item, "little difficulty" for the next item, and "some difficulty" for the hardest item (Table 1). For example, item PFC38 ("Are you able to walk at a normal speed?") is calibrated such that a "no difficulty" response aligns with a PROMIS T-score of 50 (US average). In contrast, the more difficult item PFB5r1 ("Does

Table 1. Patient-Reported Outcomes Measurement Information System (PROMIS) Item Sets and Expected Likert Responses

Physical Task	PROMIS Item ID	PROMIS Item Wording	Hypothesized Likert Response Relative to Norms (Age 60+ y)
Walking	PFC38	Are you able to walk at a normal speed?	No difficulty
	PFM23	Are you able to walk briskly for 20 min without stopping to rest?	Little difficulty
	PFB5r1	Does your health now limit you in hiking a couple of miles (3 km) on uneven surfaces, including hills?	Some difficulty
Stair ascent	PFA21	Are you able to go up- and downstairs at a normal pace?	No difficulty
	PFC32	Are you able to climb up 5 flights of stairs?	Little difficulty
	PFM21	Are you able to climb the stairs of a 10-story building without stopping?	Some difficulty
Sit to Stand	PFA15	Are you able to stand up from an armless straight chair?	No difficulty
	PFC41	Are you able to sit down in and stand up from a low, soft couch?	Little difficulty
	PFA41	Are you able to squat and get up?	Some difficulty

Abbreviations: ID, identification; PROMIS, Patient-Reported Outcomes Measurement Information System.

your health now limit you in hiking a couple of miles (3 km) on uneven surfaces, including hills?") is calibrated such that a "some difficulty" response aligns with a T-score of 50. Physical therapists might reference these PF item sets to judge meaningful changes in perceived performance.

This report seeks to improve the clinical interpretation of the PROMIS PF CAT in 3 ways: (1) by assessing convergent validity of self-reported abilities on the PROMIS PF CAT with physical performance tests in older adults; (2) by evaluating whether physical performance values predicted from PROMIS PF CAT T-scores compare with normative values for community-dwelling older adults; and (3) by determining whether sets of PROMIS PF items similar to each of the 3 timed performance tasks scale in difficulty across Likert responses. The first hypothesis was that the PROMIS PF CAT T-score and the modified Physical Performance Test (mPPT) would demonstrate convergent validity similar to previous studies (r = 0.22-0.67).[17,18,21,22] The second hypothesis was that mPPT scores predicted by PROMIS PF CAT T-scores would compare well with published normative data for selected mPPT items (ie, walking velocity, stair ascent, and 5× STS).[24-26] The third hypothesis was that for each timed performance task, a set of items from the PROMIS PF CAT item bank could be selected that scales in increasing difficulty for expected Likert responses (Table 1). These item sets would likely scale only to discriminate "no difficulty," "little difficulty," and "some difficulty" levels for each task since the sample included community-dwelling older adults and not individuals with greater functional limitations.

METHODS

Study Design

This is a cross-sectional descriptive study. The study was approved by the Institutional Review Board at George Fox University, with informed consent obtained and all rights of human participants protected.

Participants

A final sample of 45 community-dwelling older adults in rural Oregon participated. These were recruited from senior centers, local clubs (ie, Rotary, church groups), and assisted living communities. Inclusion criteria were adults 60 to 95 years of age, living independently or in assisted living environments, Mini Mental Status Exam (MMSE) score of 25 of more, stable chronic illness (eg, diabetes, osteoarthritis, peripheral vascular, or heart disease), and able to tolerate functional testing. Exclusion criteria were nonambulatory, diagnosis of neurological deficits, or acute medical problem rendering functional testing unsafe. These criteria were developed to safely allow participants with varying abilities to participate. This resulted in 5 participants being excluded for the following reasons: did not complete all testing (n = 1), nonambulatory (n = 1), and comorbidities (n = 3; peripheral neuropathy, heart disease with activity restrictions, and radicular lower extremity pain).

Participant Assessment

The participants attended 1 data collection session lasting 45 to 60 minutes. During the session, the PROMIS PF CAT and the mPPT test battery were completed, with testers blinded to the PROMIS results. The PROMIS PF CAT version 1.0 was administered to each participant on the Assessment Center Web site (assessmentcenter.net). T-scores unadjusted for age and gender were used in this analysis similar to other studies of community-dwelling older adults.[13]

The mPPT was also administered as a reliable, valid physical performance measure that assesses usual daily activities typical of older adults.[29,30] It contains a battery of 7 functional tests that include 50-m walk, stair ascent, 5× STS, picking up a penny, donning/doffing a coat, placing a book on a shelf, and turning in place. All but turning in place are timed with a stopwatch. Timed tests are converted to ordinal scores based on preestablished cutoff values. A score of 36 (range: 0-36) indicates the highest physical functioning on the mPPT.[29,30]

PROMIS PF Item Set Selection

All 124 items from the PROMIS PF CAT 1.0-item bank were considered when identifying items most similar to walking, stair ascent, or sit to stand. Using the World Health Organization ICF Activities and Participation domain coding for Walking (d450), Climbing (d4551), and Changing Basic Body Position (d410),[31] 37 PROMIS PF items were initially identified (by authors J.H. and R.J.) as fitting 1 of these 3 ICF activity codes. Coding ascribed to these 37 matched exactly with previous work by Tucker et al,[28] who used an item-level content analysis mapping method to ascribe World Health Organization–ICF codes to all 124 items in the PROMIS PF CAT 1.0-item bank. Three PF items were then selected for each performance task (ie, walking, stair ascent, or sit to stand) that were most similar to each task *and* that best discriminated increasing difficulty such that an average older adult would likely self-report "no difficulty" for the easiest item, "little difficulty" for the next item, and "some difficulty" for the hardest item. This was achieved through visual side-by-side comparison of T-score difficulty ranges, based on item model parameter calibration, for items in each of the 3 ICF activity codes.

Statistical Analysis

Descriptive statistics were calculated for the sample. SPSS (v25, IBM Corporation) software was used to calculate all correlations and regression equations. Each hypothesis was analyzed as follows.

Hypothesis 1

Convergent validity was tested using Pearson correlations between mPPT measures (total score and each timed component item) and PROMIS PF CAT T-scores. Expecting correlations of r = 0.4 to 0.5,[17,18,21,22] a sample size of 44 was necessary to achieve sufficient power (β = .2 and

α = .05) for an *r* value of 0.41 or higher. For the 6 timed items that make up the mPPT test, a Bonferroni correction was applied to the α level to decrease the chance of spurious findings (*P* = .05/6 = .0083). For mPPT variables, walking velocity, stair ascent, and 5× STS (significantly correlated with the PROMIS PF CAT T-scores) equations of the line were calculated and used to estimate mPPT values from the T-scores. The standard error of the estimate was used to evaluate fit and reflect the range of possible mPPT test item scores (eg, range of times in seconds). Note that all significant correlations met assumptions of normality (assessed by normal probability plots) and homoscedasticity (see residual plots in Appendix 1). Two outliers were identified using Mahalanobis and Cooks distance. However, on review and visual inspection of scatter plots (Appendix 1), their identification as outliers occurred because they performed better than the group, so their values were retained. Spearman correlations were applied to demographic (age and gender) and health covariates (number of comorbidities, medications, and falls; body mass index; Mini-Mental State Score) with the PROMIS PF CAT T-score. The intent of the Spearman correlations was to demonstrate the potential influence of these covariates to modify hypothesis 1.

Hypothesis 2

Published norms for walking velocity,[26] stair ascent,[25] and 5× STS[24] in adults 60 years of age and older were compared with regression-predicted mPPT timed component performance. Although a significant correlation was also found for picking up a penny with PROMIS PF CAT T-score, no published normative data are available for the task and no adequately similar PROMIS items were identified from the item bank.

Hypothesis 3

The PROMIS PF item response theory model parameters for the selected PF items were analyzed for their ability to discriminate increasing difficulty level (Likert responses) alongside regression-predicted mPPT timed component performance and normative data. Item response theory model parameters were obtained from PROMIS developers collaborating on this project (author M.B.). The model parameters calibrate each PF item, identifying the T-score demarcation points between difficulty responses (ie, "no difficulty," "little difficulty," "some difficulty," "much difficulty," and "unable to do"). Using these T-score demarcation points, the final 9 items (3 for each performance task) were aligned to predicted mPPT values and normative data and then displayed in a table and used to generate a visual "Snapshot" for improved clinical interpretation of PROMIS PF CAT T-scores.

Table 2. Descriptive Statistics of the Sample

	n	Average (SD)/Median[a] or Frequency (%)	Range
Age, y	45	77.1 (4.6)	65-84
Females	27	60%	
Height, cm	45	167.1 (10.9)	151.1-188.0
Weight, kg	45	74.0 (17.6)	37.7-125.0
BMI, kg/m^2	45	26.2 (5.0)	14.7-39.5
Comorbidities	45	3[a]	0-6
Medications	45	3[a]	0-18
Falls within past year	45		
0	35	77.8%	
1	7	15.6%	
2	3	6.6%	
Mini-Mental score	45	28.4 (1.5)	25-30
mPPT Total score	45	29.1 (3.7)	20-35
5× STS, s	45	14.5 (3.5)	9-27
Stair ascent, s	45	7.4 (2.0)	5-14
Book, s	45	3.0 (1.4)	1-11
Coat, s	45	13.5 (7.9)	7-60
Penny, s	45	3.7 (1.9)	2-13
50-m walk, s	45	12.4 (3.1)	8-22
50-m walk, m/s	45	1.29 (0.29)	0.7-2.0
PROMIS PF Scale	45	49.4 (5.0)	41-62

Abbreviations: BMI, body mass index; mPPT, modified Physical Performance Test; 5× STS, 5 times sit to stand; PROMIS PF, Patient-Reported Outcomes Measurement Information System Physical Function.
[a]Median value.

RESULTS

The participants were on average 77.1 (4-6) years of age and had a mean body mass index of 26.2 (5-0) (Table 2). The participants' mean MMSE score was 28.3 (1-5), the median number of comorbidities was 3 (range: 0-6), and the median number of medications was 3 (range: 0-18). Ten participants (21.7%) experienced 1 or more falls to the ground in the last year. The average mPPT score was 29.1 (3-7). The average PROMIS PF CAT T-score was 49.4 (5-0) (near the average 50 for the US population). Body mass index, MMSE score, and number of falls in the last year were significantly correlated to PROMIS PF CAT T-scores (Table 3).

The PROMIS PF CAT T-score showed significant univariate correlations at or above *r* value of 0.47, with the mPPT total score and 4 of the 6 timed items that make up the mPPT test battery (Table 3). The highest correlation was between walking velocity and PROMIS PF CAT ($r = 0.64$, $P < .001$). Placing a book overhead ($r = -0.32$, $P = .04$) and timed donning/doffing a coat ($r = -0.28$, $P = .07$) were not significant after applying the Bonferroni correction.

Using the regression equations, it was possible to link PROMIS PF CAT T-scores to predicted mPPT values and published norms for walking velocity, stair ascent, and 5× STS (Table 4). Regarding individual mPPT items, a T-score of 50 corresponded to a walking velocity of 1.31 m/s, a

Table 3. Univariate Correlations of Patient-Reported Outcomes Measurement Information System Physical Function T-Scores With Participant Descriptors and Physical Performance Tests

Variable	*r*	Standard Error of the Estimate (% Average)	*P*	Equation of the Line
Demographic				
Age	−0.10		**.49**	
BMI	−0.37		**.01**	
Mental state				
Mini-Mental State Examination	0.37		**.01**	
Covariates				
Number of medications	−0.15[a]		**.33**	
Number of comorbidities	−0.27[a]		**.07**	
Number of falls	−0.36[a]		**.02**	
Modified Physical Performance Test				
Overall score	0.59	3.00 (10.1%)	<**.01**	7.87 + T-score (0.43)
Individual item scores[b]				
5× STS, s	−0.47	3.11 (21.4%)	<**.001**	30.59 + T-score (−0.33)
Stair ascent, s	−0.54	1.71 (23.1%)	<**.001**	17.82 + T-score (−0.21)
Book, s	−0.32		.04	ns
Coat, s	−0.28		.07	ns
Penny, s	−0.47	1.76 (47.6%)	<**.001**	12.6 + T-score (−0.18)
Walking velocity, m/s	0.64	0.22 (17.0%)	<**.001**	T-score (0.04) –0.54

Abbreviations: BMI, body mass index; 5× STS, 5 times sit to stand; ns, not significant.
[a]Ordinal data *r* values calculated using a Spearman ρ statistic; all other *r* values calculated using Pearson correlation.
[b]Values in boldface for individual item scores indicate *P* values < .0083.

timed stair ascent of 7.30 s, and a 5× STS of 14.29 s, all falling within published norms (bolded in Table 4). A PROMIS PF CAT T-score of 44 to 52 spanned the normal range for walking speed, 49 to 61 for stair ascent, and 49 to 58 for 5× STS. The lowest T-score was 41, which corresponded to a walking velocity of 0.98 m/s, a timed stair ascent of 9.19 s, and a 5× STS of 17.23 s. These values were all slower than published norms for older adults. The highest T-score of 62 corresponded to a walking velocity of 1.76 m/s, a timed stair ascent of 4.77 s, and a 5× STS of 10.38 s, all faster than published norms.

Referencing a T-score of 50 as representative of normative for community-dwelling older adults, the selected PROMIS PF items most similar to stair ascent and walking were largely able to discriminate increasing levels of difficulty. However, the PF items most similar to sit to stand only discriminated levels of function for T-scores below 50. As is shown in Table 4, a T-score of 50 aligned best with hypothesized self-report responses for stair ascent—"no difficulty" for PFA21, "little difficulty" for PFC32, and "some difficulty" for PFM21. A T-score of 50 aligned with "no difficulty" for walking item PFC38 and also for PFM23 and with "little difficulty" for PFB5r1. For sit to stand, a T-score of 50 aligned with "no difficulty" for item PFA15, as hypothesized. However, 50 also aligned with "no difficulty" for PFA41 and PFC41 (originally hypothesized to align with responses of "little difficulty" and "some difficulty," respectively). The 9 PF items are also displayed in a stacked column chart (ie, "Snapshot") for clinical interpretation across *all* PROMIS PF CAT T-scores (Appendix 2).

DISCUSSION

These data advance interpretation of PROMIS PF CAT T-scores for clinical decision making for individuals with T-scores between 41 and 62. The participants' perceived ability on the PROMIS PF CAT significantly correlated with the mPPT and 4 of the 6 timed mPPT component items tested, demonstrating convergent validity (hypothesis 1) (Table 3). As expected, the standard error of the estimates ranged from 17.0% to 23.1%, suggesting that the PROMIS PF CAT T-scores are useful for identifying a majority of individuals (68.1%) within a range of their predicted score (Table 4). Also as expected, the PROMIS PF CAT T-score of 50 compared positively with published normative data for performance tests (hypothesis 2) (Table 4, values in boldface). The items selected from the PF item bank (Table 1) showed some utility clinically for stair ascent and walking to help providers understand changes in

Table 4. Physical Performance Tests, Published Norms, and PROMIS Physical Function Item Sets Linked to PROMIS T-Score

<table>
<tr><th>PROMIS T-score</th><th>PREDICTED mPPT Total Score</th><th>PREDICTED Walking Velocity[a], m/s</th><th>PFC38 Walk at normal speed</th><th>PFM23 Walk briskly 20 minutes, no rest</th><th>PFB5r1 Hike 2 miles, uneven surfaces</th><th>PREDICTED Stair Ascent[b],s</th><th>PFA21 Up and down stairs, normal pace</th><th>PFC32 Climb 5 flights stairs</th><th>PFM21 Climb 10 stories stairs, no rest</th><th>PREDICTED Five Times Sit-to-Stand[c],s</th><th>PFA15 Stand from armless chair</th><th>PFC41 Sit then stand from low couch</th><th>PFA41 Squat then get up</th></tr>
<tr><td>41</td><td>26</td><td>0.98</td><td rowspan="2">Little Difficulty</td><td rowspan="3">Some Difficulty</td><td rowspan="2">Much Difficulty</td><td>9.19</td><td rowspan="5">Little Difficulty</td><td rowspan="5">Some Difficulty</td><td rowspan="5">Much Difficulty</td><td>17.23</td><td rowspan="2">Little Difficulty</td><td rowspan="6">Little Difficulty</td><td rowspan="9">Little Difficulty</td></tr>
<tr><td>42</td><td>26</td><td>1.02</td><td>8.98</td><td>16.90</td></tr>
<tr><td>43</td><td>26</td><td>1.05</td><td rowspan="20">No Difficulty</td><td rowspan="5">Some Difficulty</td><td>8.77</td><td>16.57</td><td rowspan="20">No Difficulty</td></tr>
<tr><td>44</td><td>27</td><td>1.09</td><td rowspan="6">Little Difficulty</td><td>8.56</td><td>16.25</td></tr>
<tr><td>45</td><td>27</td><td>1.13</td><td>8.35</td><td>15.92</td></tr>
<tr><td>46</td><td>28</td><td>1.17</td><td>8.14</td><td rowspan="17">No Difficulty</td><td rowspan="7">Little Difficulty</td><td rowspan="7">Some Difficulty</td><td>15.60</td></tr>
<tr><td>47</td><td>28</td><td>1.20</td><td>7.93</td><td>15.27</td><td rowspan="16">No Difficulty</td></tr>
<tr><td>48</td><td>29</td><td>1.24</td><td rowspan="5">Little Difficulty</td><td>7.72</td><td>14.94</td></tr>
<tr><td>49</td><td>29</td><td>1.28</td><td>7.51</td><td>14.62</td></tr>
<tr><td>50</td><td>29</td><td>1.31</td><td rowspan="13">No Difficulty</td><td>7.30</td><td>14.29</td><td rowspan="13">No Difficulty</td></tr>
<tr><td>51</td><td>30</td><td>1.35</td><td>7.09</td><td>13.97</td></tr>
<tr><td>52</td><td>30</td><td>1.39</td><td>6.88</td><td>13.64</td></tr>
<tr><td>53</td><td>31</td><td>1.42</td><td rowspan="10">No Difficulty</td><td>6.67</td><td rowspan="10">No Difficulty</td><td rowspan="8">Little Difficulty</td><td>13.31</td></tr>
<tr><td>54</td><td>31</td><td>1.46</td><td>6.46</td><td>12.99</td></tr>
<tr><td>55</td><td>32</td><td>1.50</td><td>6.25</td><td>12.66</td></tr>
<tr><td>56</td><td>32</td><td>1.54</td><td>6.04</td><td>12.34</td></tr>
<tr><td>57</td><td>32</td><td>1.57</td><td>5.83</td><td>12.01</td></tr>
<tr><td>58</td><td>33</td><td>1.61</td><td>5.62</td><td>11.68</td></tr>
<tr><td>59</td><td>33</td><td>1.65</td><td>5.40</td><td>11.36</td></tr>
<tr><td>60</td><td>34</td><td>1.68</td><td>5.19</td><td>11.03</td></tr>
<tr><td>61</td><td>34</td><td>1.72</td><td>4.98</td><td rowspan="2">No Difficulty</td><td>10.71</td></tr>
<tr><td>62</td><td>35</td><td>1.76</td><td>4.77</td><td>10.38</td></tr>
</table>

Abbreviations: mPPT, modified Physical Performance Test; PROMIS, Patient-Reported Outcomes Information System.
[a]Walking velocity norms (95% confidence interval: ***1.07-1.41 m/s***) for adults aged 60 to 79 years.[26]
[b]Stair ascent norms (range of means: ***4.92-7.56 s***) for adults aged 65 years and older.[25]
[c]5x STS norms (95% confidence interval: ***11.4-14.8 s***) for adults aged 60 to 89 years.[24]

self-reported function; however, the available items for sit to stand were less useful (hypothesis 3).

The level of convergent validity between a patient-reported outcomes measure (ie, PROMIS PF CAT) and physical performance testing (ie, mPPT) is consistent with past studies of community-dwelling older adults (r value range = 0.56-0.65)[16-18] and individuals after surgery (r value range = 0.40-0.67).[19-23] The same range of convergence was shown in studies using the same or similar measures for community-dwelling older adults (self-reported "need for help" on 4 activities of daily living, stair ascent, and distance walking vs Short Physical Performance Battery; r = 0.65),[17] patients post–hip fracture (lower extremity measure vs mPPT total score; r = 0.55 vs timed stair ascent; r = 0.47),[20] and patients with obstructive pulmonary disease (PROMIS PF CAT 1.0 vs 6-minute walk test; r = 0.57).[18] Slightly higher convergence (r = 0.65-0.67) was noted by Latham et al[23] in patients post–hip fracture comparing the Activity Measure for Post-Acute Care Physical Mobility scale with the Short Physical Performance Battery, 6-minute walk test, and self-selected walking velocity. In short, despite using new state-of-the-art procedures to develop the PROMIS PF CAT scale (ie, item response theory and computer-adaptive algorithms), the convergence of self-report and measured performance is consistent with current and previous studies.

The standard errors for estimating timed performance on the 3 mPPT tasks based on T-scores (Table 3) suggest

the PROMIS PF CAT scale is useful for narrowing the expected range of actual performance, which can then be confirmed with focused physical performance testing as necessary. The standard errors of the estimate ranged from 17.0% to 23.1%. This suggests that predictions based on PROMIS PF CAT lie within the range of actual mPPT scores from this sample (Table 2) approximately 68% of the time. These ranges are reflective of the correlations between the PROMIS PF CAT and performance tests for this study and past studies.[16-18] Although the correlations are not strong enough to accurately determine individual performance, they provide a starting point for physical testing or creating goals. Future studies could consider larger and/or more homogeneous samples to examine whether potentially higher correlations improve accuracy and/or narrow the error range.[23]

The normative values from published articles compared well with the PROMIS PF CAT T-score of 50. As shown in Table 4, a T-score of 50 (average for the general US population) predicts a walking velocity of 1.31 m/s, with a range of 1.09 to 1.53 m/s. This predicted result is centered over the 95% confidence interval lower and upper limits (1.07-1.41 m/s) described in a descriptive meta-analysis by Bohannon et al[26] as normal for men and women aged 60 to 79 years. A T-score of 50 predicts a stair ascent time of 7.30 s with a range of 5.59 to 9.01 s, overlapping the range of norms (4.92 to 7.56 s) from the systematic review by Nightingale et al[25] for adults aged 65 years and older. A T-score of 50 predicts a 5× STS time of 14.29 s, with a range of 11.18 to 17.40 s, overlapping with the 95% confidence interval reported for adults aged 60 to 89 years (11.4-14.8 s) in a meta-analysis by Bohannon.[24] Most recently, Bohannon and colleagues[24,32] described lower 5× STS times of 10.1 (2-6) for healthy mobile adults aged 74 to 97 years, pointing out the variability of 5× STS times reported in the literature. Nevertheless, our data for 45 community-dwelling older adults overlap with these published normative data sets, supporting use of the PROMIS PF CAT T-scores to determine a patient's expected range of performance based on his or her self-report.

The 9 PROMIS PF items that best scaled in increasing difficulty across T-scores, predicted timed performance, and normative data are identified (Table 4) and then visually displayed (Appendix 2). Items most similar to stair ascent scaled best into 3 distinct difficulty responses (from "no difficulty" to "some difficulty") at a T-score of 50 (Table 4). This suggests that an individual's responses on these 3 stair ascent items would likely discriminate level of functioning in a clinically meaningful way. The PF items most similar to walking discriminated between "no difficulty" and "little difficulty" at a T-score of 50. However, the most difficult item available in the item bank (PFB5r1 - "Does your health now limit you in hiking a couple of miles (3 km) on uneven surfaces, including hills?") did align with "some difficulty" at a T-score of 47, only slightly below the US average 50. This likely means that PROMIS PF walking items are less meaningful in discriminating higher-functioning individuals, even when the PF walking item seems difficult at face value (eg, PFB5r1). The PF items most similar to sit to stand were the least effective at discriminating level of functioning. Even the most difficult sit-to-stand related item available in the PF item bank (PFC41—"Are you able to sit down in and stand up from a low, soft couch?") aligned only with a T-score of 49 to demarcate a change from "little difficulty" up to "no difficulty." From a biomechanical perspective, the sit-to-stand task is more difficult than either walking or stair ascent,[33,34] but perhaps for a frequently practiced daily task such as sit to stand, self-perception of ability is a more relevant factor. Future studies may consider new PF items associated with less-practiced tasks with higher demands for endurance or balance, and/or evaluate why self-report of sit to stand does not discriminate higher functional abilities.

These results improve interpretation of the PROMIS PF CAT scale. Table 4 and the "Snapshot" (Appendix 2) are graphical representations of the correspondence between PROMIS PF CAT T-score, predicted mPPT physical performance, normative data for older adults, and expected PF item Likert responses for difficulty. Using Table 4 or the "Snapshot," a physical therapist might choose to start an initial evaluation and set goals based on a quickly obtained PROMIS PF CAT T-score. For example, a T-score of 41 suggests that the expected responses are "much difficulty" for PFB5r1 (hiking 2 miles) and PFM21 (10-story stair ascent), and "some difficulty" for PFM23 (brisk 20-minute walk) and PFC32 (5-flight stair ascent) (Appendix 2). Using predicted scores (Table 4) and standard errors (Table 3), this individual's present physical performance abilities would be approximated at a walking velocity between 0.76 and 1.20 m/s, stair ascent time between 7.48 and 10.90 s (1 flight), and 5× STS between 14.12 and 20.34 s. Therefore, evaluation and initial goal setting might focus on speed and endurance in walking and stair ascent. In contrast, an individual with a T-score of more than 53 is likely to have "no difficulty" with 8 of the 9 selected PROMIS PF items. Present physical performance would be approximated by a walking velocity of 1.20 to 1.64 m/s, stair ascent time of 4.96 to 8.38 s, and 5× STS of 10.20 to 16.42 s. For this individual, more difficult functional tasks should be considered immediately and higher-level goals identified.

Although these data are useful for community-dwelling older adults, the sample size and range of abilities of the participants may not generalize to clinical populations. Individuals with PROMIS PF CAT T-scores higher or lower than the participants in this study (>62 or <41) may show different associations. For example, the 5× STS task may show better ability to differentiate PROMIS PF CAT T-scores below 41. Therefore, these data may apply most directly to community-dwelling older adults participating in exercise or prevention programs. Because the PROMIS PF CAT scale achieves a T-score in relatively few questions (as few as 4) and can be applied across diagnoses, this scale

may be ideal for large-scale research related to prevention and wellness in older adult populations.

Variables such as body mass index, MMSE, and number of falls showed significant correlations with the PROMIS PF CAT scale (Table 3) and are likely useful to take into account in future multivariate analyses, enhancing predictive capabilities. Although the sample size was powered for *r* values of 0.41, a larger sample could verify the strength of the associations identified in this study.[17,23] The use of age-adjusted scores or percentiles was not applied in this analysis. Because this study's sample average T-score (49.4) matches the US population average and normative data are also from community-dwelling older adults, the unadjusted US average is referenced throughout. However, other studies may consider using adjusted percentile scores and including clinical populations with greater dependence.[13,14] Future studies might also consider benchmarking PROMIS PF CAT T-score thresholds for key clinical decisions (eg, discharge disposition from inpatient settings, need for in-home services), further increasing the utility of T-scores.

CONCLUSIONS

The PROMIS PF CAT T-scores showed expected convergent validity with physical performance measures. The correlations enable estimates to be generated between ranges of physical performance and the PROMIS PF CAT T-score. While the findings are not predictive of individual performance, they improve clinical interpretation by estimating a range of performance based on PROMIS PF CAT T-scores for walking, stair ascent, and sit to stand. Meaningfulness of PROMIS PF CAT T-scores was strengthened by correspondence of predicted mPPT test scores for walking velocity, stair ascent, and 5× STS with normative data for older adults. Individual PROMIS PF item responses showed some ability to discriminate levels of difficulty for specific tasks. It is anticipated that these data will be useful for optimizing clinical decision making in evaluation and initial goal setting for community-dwelling older adults.

ACKNOWLEDGMENTS

The authors are grateful for the contribution of Tomoko Iwanaga, Christine Kieu, Keegan Lerma, Chelan Murasaki, and Tawny Souza to data collection as part of their doctoral studies at George Fox University, Newberg, Oregon.

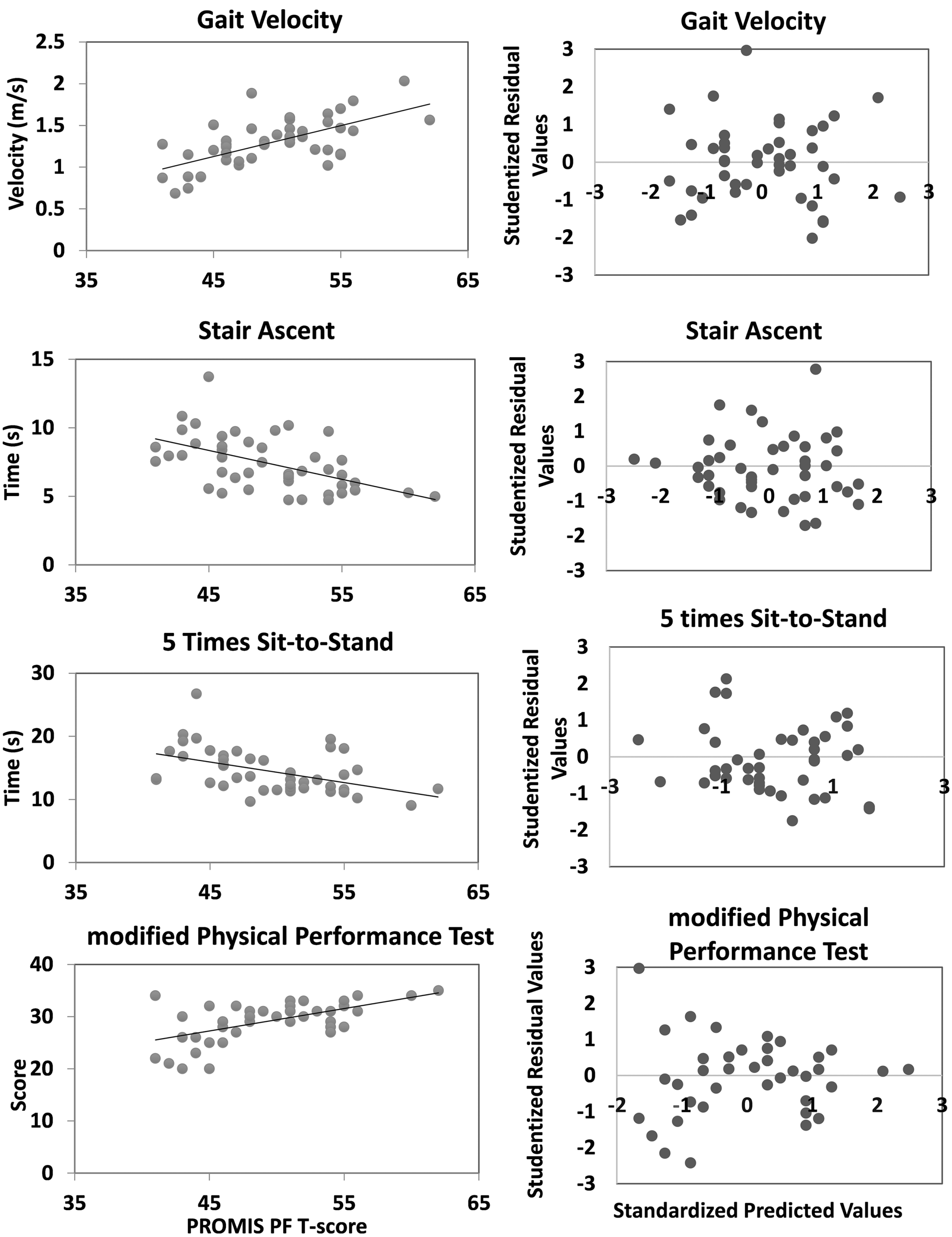
REFERENCES

1. Hung M, Baumhauer JF, Brodsky JW, et al. Psychometric comparison of the PROMIS Physical Function CAT with the FAAM and FFI for measuring patient-reported outcomes. *Foot Ankle Int*. 2014;35(6):592-599.
2. Hung M, Baumhauer JF, Latt LD, Saltzman CL, Soohoo NF, Hunt KJ. Validation of PROMIS® Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res*. 2013;471(11):3466-3474.
3. Papuga MO, Beck CA, Kates SL, Schwarz EM, Maloney MD. Validation of GAITRite and PROMIS as high-throughput physical function outcome measures following ACL reconstruction. *J Orthop Res*. 2014;32(6):793-801.
4. Papuga MO, Mesfin A, Molinari R, Rubery PT. Correlation of PROMIS Physical Function and Pain CAT instruments with Oswestry Disability Index and Neck Disability Index in spine patients. *Spine*. 2016;41(14):1153-1159.
5. Brodke DS, Goz V, Voss MW, Lawrence BD, Spiker WR, Hung M. PROMIS PF CAT outperforms the ODI and SF-36 physical function domain in spine patients. *Spine*. 2017;42(12):921-929.
6. Hung M, Franklin JD, Hon SD, Cheng C, Conrad J, Saltzman CL. Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. *Foot Ankle In*. 2014;35(1):1-7.
7. Papuga MO, Dasilva C, McIntyre A, Mitten D, Kates SL, Baumhauer JF. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. *Health Syst*. 2017;7(2):1-12.
8. Baumhauer JF. Patient-reported outcomes—are they living up to their potential? *N Engl J Med*. 2017;377(1):6-9.
9. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. *Arch Phys Med Rehabil*. 2011;92 (10 suppl):S12-S19.
10. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, 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(7):873-880.
11. Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a common metric for physical function: linking the HAQ-DI and SF-36 PF subscale to PROMIS((R)) Physical Function. *J Gen Intern Med*. 2015;30(10):1517-1523.
12. Cella D, Choi S, Garcia S, et al. Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Qual Life Res*. 2014;23(10):2651-2661.
13. Balachandran A, Martins MM, De Faveri FG, Alan O, Cetinkaya F, Signorile JF. Functional strength training: seated machine vs standing cable training to improve physical function in elderly. *Exp Gerontol*. 2016;82:131-138.
14. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol*. 2014;67(5):516-526.
15. Sheean AJ, Schmitz MR, Ward CL, et al. Assessment of disability related to femoroacetabular impingement syndrome by use of the Patient-Reported Outcome Measure Information System (PROMIS) and objective measures of physical performance. *Am J Sports Med*. 2017;45(11):2476-2482.
16. Adegoke BO, Babatunde FO, Oyeyemi AL. Pain, balance, self-reported function and physical function in individuals with knee osteoarthritis. *Physiother Theory Pract*. 2012;28(1):32-40.
17. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49(2):M85-M94.
18. Irwin DE, Atwood CA Jr, Hays RD, et al. Correlation of PROMIS scales and clinical measures among chronic obstructive pulmonary disease patients with and without exacerbations. *Qual Life Res*. 2015;24(4):999-1009.
19. Swinkels A, Allain TJ. Physical performance tests, self-reported outcomes, and accidental falls before and after total knee arthroplasty: an exploratory study. *Physiother Theory Pract*. 2013;29(6):432-442.
20. Briggs R, Houck J, Drummond M, Fritz J, LaStayo P, Marcus R. Asymmetries identified in sit-to-stand task explain physical function after hip fracture *J Geriatr Phys Ther*. 2018;41(4):210-217.
21. Farag I, Sherrington C, Kamper SJ, et al. Measures of physical functioning after hip fracture: construct validity and responsiveness of performance-based and self-reported measures. *Age Ageing*. 2012;41(5):659-664.
22. Jaglal S, Lakhani Z, Schatzker J. Reliability, validity, and responsiveness of the lower extremity measure for patients with a hip fracture. *J Bone Joint Surg Am*. 2000;82-A(7):955-962.
23. Latham NK, Mehta V, Nguyen AM, et al. Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? *Arch Phys Med Rehabil*. 2008;89(11):2146-2155.
24. Bohannon RW. Reference values for the five-repetition sit-to-stand test: a descriptive meta-analysis of data from elders. *Percept Mot Skills*. 2006;103(1):215-222.
25. Nightingale EJ, Pourkazemi F, Hiller CE. Systematic review of timed stair tests. *J Rehabil Res Dev*. 2014;51(3):335-350.
26. Bohannon RW, Williams Andrews A. Normal walking speed: a descriptive meta-analysis. *Physiother*. 2011;97(3):182-189.
27. Tucker CA, Cieza A, Riley AW, et al. Concept analysis of the Patient Reported Outcomes Measurement Information System (PROMIS) and the International Classification of Functioning, Disability and Health (ICF). *Qual Life Res*. 2014;23(6):1677-1686.
28. Tucker CA, Escorpizo R, Cieza A, et al. Mapping the content of the Patient-Reported Outcomes Measurement Information System (PROMIS(R)) using

the International Classification of Functioning, Health and Disability. *Qual Life Res*. 2014;23(9):2431-2438.
29. Brown M, Sinacore DR, Binder EF, Kohrt WM. Physical and performance measures for the identification of mild to moderate frailty. *J Gerontol A Biol Sci Med Sci*. 2000;55(6):M350-M355.
30. Brown M, Sinacore DR, Ehsani AA, Binder EF, Holloszy JO, Kohrt WM. Low-intensity exercise as a modifier of physical frailty in older adults. *Arch Phys Med Rehabil*. 2000;81(7):960-965.
31. World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: World Health Organization; 2001.
32. Bohannon RW, Wolfson LI, White WB. Timed mobility: description of measurement, performance, and dimensionality among older adults. *Disabil Rehabil*. 2018;40(17):2011-2014.
33. Kneiss JA, Hilton TN, Tome J, Houck JR. Weight-bearing asymmetry in individuals post-hip fracture during the sit to stand task. *Clin Biomech*. 2015;30(1):14-21.
34. Nadeau S, McFadyen BJ, Malouin F. Frontal and sagittal plane analyses of the stair climbing task in healthy adults aged over 40 years: what are the challenges compared to level walking? *Clin Biomech*. 2003;18(10):950-959.

APPENDIX 1

Residual Plots Demonstrating Homoscedasticity (Homogeneity of Variance) in Regression Plots for Walking Velocity, Stair Ascent, 5 Times Sit to Stand, and Modified Physical Performance Test Total Score



APPENDIX 2

"Snapshot" Benchmarking-Selected PROMIS Physical Function Items Against Overall T-Score