A Physical Function Test for Use in the Intensive Care Unit: Validity, Responsiveness, and Predictive Utility of the Physical Function ICU Test (Scored)

Linda Denehy, Natalie A. de Morton, Elizabeth H. Skinner, Lara Edbrooke, Kimberley Haines, Stephen Warrillow, Sue Berney

Background. Several tests have recently been developed to measure changes in patient strength and functional outcomes in the intensive care unit (ICU). The original Physical Function ICU Test (PFIT) demonstrates reliability and sensitivity.

Objective. The aims of this study were to further develop the original PFIT, to derive an interval score (the PFIT-s), and to test the clinimetric properties of the PFIT-s.

Design. A nested cohort study was conducted.

Methods. One hundred forty-four and 116 participants performed the PFIT at ICU admission and discharge, respectively. Original test components were modified using principal component analysis. Rasch analysis examined the unidimensionality of the PFIT, and an interval score was derived. Correlations tested validity, and multiple regression analyses investigated predictive ability. Responsiveness was assessed using the effect size index (ESI), and the minimal clinically important difference (MCID) was calculated.

Results. The shoulder lift component was removed. Unidimensionality of combined admission and discharge PFIT-s scores was confirmed. The PFIT-s displayed moderate convergent validity with the Timed "Up & Go" Test ($r=-.60$), the Six-Minute Walk Test ($r=.41$), and the Medical Research Council (MRC) sum score (rho=.49). The ESI of the PFIT-s was 0.82, and the MCID was 1.5 points (interval scale range=0–10). A higher admission PFIT-s score was predictive of: an MRC score of ≥48, increased likelihood of discharge home, reduced likelihood of discharge to inpatient rehabilitation, and reduced acute care hospital length of stay.

Limitations. Scoring of sit-to-stand assistance required is subjective, and cadence cutpoints used may not be generalizable.

Conclusions. The PFIT-s is a safe and inexpensive test of physical function with high clinical utility. It is valid, responsive to change, and predictive of key outcomes. It is recommended that the PFIT-s be adopted to test physical function in the ICU.

L. Denehy, BAppSc(Physio), Grad DipPhysio(Cardiothoracic), PhD, Department of Physiotherapy, The University of Melbourne, Parkville, Victoria, Australia 3000. Address all correspondence to Ms Denehy at: l.denehy@unimelb.edu.au.

N.A. de Morton, BAppSci(Physio), PhD, Department of Physiotherapy, Donvale Rehabilitation Hospital & Peninsula Private Hospital, Ramsay Health Victoria, Victoria, Australia.

E.H. Skinner, BPhysiotherapy (Hons), PhD, Department of Physiotherapy, The University of Melbourne; Allied Health Research Unit, Southern Health, Victoria, Australia; and Physiotherapy Department, Austin Health, Heidelberg, Victoria, Australia.

L. Edbrooke, BAppSci(Physio), Grad DipEpidemiolBiostats, Department of Physiotherapy, The University of Melbourne.

K. Haines, BHSc(Physiotherapy), Physiotherapy Department, Austin Health.

S. Warrillow, MBBS, FCICM, FRACP, Department of Intensive Care, Austin Health.

S. Berney, BPhysio, MPhysio, PhD, Physiotherapy Department, Austin Health.

[Denehy L, de Morton NA, Skinner EH, et al. A physical function test for use in the intensive care unit: validity, responsiveness, and predictive utility of the Physical Function ICU Test (scored). *Phys Ther.* 2013;93:xxx–xxx.]

Published Ahead of Print: July 25, 2013
Accepted: July 19, 2013
Submitted: October 15, 2012

With a growing worldwide interest in early rehabilitation in the intensive care unit (ICU),[1–3] there is now a challenge to develop sensitive and appropriate methods of measuring change in patient strength and functional outcomes. Measures used in other patient populations and clinical settings to prescribe and evaluate the effects of exercise programs such as the Six-Minute Walk Test (6MWT) or Timed "Up & Go" Test (TUG) are impractical in the ICU environment. These tests require space to perform and may require management of several drips, drains, and oxygen delivery systems while the patient is walking and turning that render the test difficult to carry out.

Several recent articles have described tests that have been specifically designed for use in the ICU phase of the patient continuum of care.[4–6] However, few tests have reported reliability and sensitivity in this patient population. Skinner et al[5] reported excellent reliability for the Physical Function ICU Test (PFIT) using the intraclass correlation coefficient (ICC; range=.996–1.00) and sensitivity to change (mean increase in cadence, knee extension strength, and shoulder flexion strength) in a small sample of patients posttracheostomy.[5] A second test, the Functional Status Score for the Intensive Care Unit (FSS-ICU) test, has high clinical utility and demonstrated a small responsiveness to change in 101 patients at a long-term acute care facility (effect size=0.25).[7] The original PFIT had low clinical utility in its original form, with each test component reported separately. Given the burgeoning volume of research related to strength and mobility in the ICU, there is an urgent need for objective functional tests with robust clinimetric properties and high clinical utility. Scoring the original PFIT offers scope for improved clinical utility and use in research to compare function and response to intervention in ICU patient populations. Additionally, developing an interval score from the ordinal "rank" score may provide added advantages in relation to interpretation and precision in research.[8]

The primary aim of this research was to use Rasch analysis to assess the fit of the ordinal PFIT items and the unidimensionality of the test. If found to be unidimensional, we aimed to transform the ordinal scoring system to an interval scoring system. The secondary aim was to assess the following clinimetric properties of the newly developed interval-scored Physical Function ICU Test (PFIT-s): validity, responsiveness, minimal clinically important difference (MCID), and predictive ability.

Method

Study Design

This study was nested within our larger randomized controlled trial (RCT)[9] measuring the effectiveness of exercise rehabilitation in survivors of the ICU.

Participants

One hundred fifty participants recruited to our RCT were assessed. The study protocol for the larger RCT was published previously, and inclusion and exclusion criteria are detailed in the protocol.[9] All participants provided written informed consent initially or continuation of consent from initial next of kin consent.

For the nested study, we calculated that a sample size of 100 would provide 95% confidence within ±0.5 logits.[10] Therefore, Rasch analysis was conducted on 2 independent samples: ICU admission and discharge PFIT data. All available data were used in the Rasch analyses (n=144 at ICU admission and n=116 at ICU discharge). Paired data were used for clinimetric calculations (n=116).

Procedure

The original test was previously reported,[5] and original components are given in Table 1. The PFIT was administered by trained physical therapists who were between 4 and 6 years from graduation, and 5 different physical therapists performed the tests during data collection. The PFIT was measured at (or near) day 5 post-ICU admission (recruitment time point in RCT) if participants were able to follow 3 of 5 simple commands to measure wakefulness[11] and at ICU discharge. Participants who were unable to perform the test between trial recruitment and 10 days of ICU admission were scored as zero as this was considered an indication of the effect of illness severity on functional capacity. The test components requiring effort (marching on the spot and bilateral shoulder lifts) were performed using the Borg Scale of Perceived Exertion,[12] where patients were asked to work to a Borg scale score between 3 and 4 on the modified scale. This Borg scale score represents "moderate" to "somewhat hard" levels of exertion and was used to permit prescription of the same *relative* intensity across participants. The procedure for the test was described previously.[5]

All tests were performed once the participants were slide transferred from bed to sit in a chair. The sit-to-stand component from the chair was performed first and then marching in place once standing, followed by the 2 strength tests once seated again in the chair. Instructions and encouragement were standardized throughout the test. The number of assistants to aid standing from sitting is part of the scoring system of the test; no other aids were used in this component. Walking frames were used if

Table 1.
Components of the Original and Modified Versions of the Physical Function ICU Test (PFIT)

Original 5-Component PFIT	New 4-Item PFIT
Assistance (sit to stand)[a]	Assistance (sit to stand)[a]
Cadence (steps/min)[b]	Cadence (steps/min)[b]
Shoulder (flexion strength)[c]	Shoulder (flexion strength)[c]
Knee (extension strength)[c]	Knee (extension strength)[c]
Bilateral shoulder lifts (lifts/min)	

[a] Sit-to-stand assistance (0, 1, or 2 people needed).
[b] Calculated on maximal marching on the spot duration and number of steps.
[c] Greatest of left and right using the Oxford grading system (muscle strength recorded as: 0=no contraction, 1=visible/palpable muscle contraction, 2=movement across gravity, 3=movement against gravity, 4=movement against gravity with some resistance, or 5=movement against gravity with full resistance.

needed to provide support once the participant was standing, but no other assistance was offered when marching in place.

The PFIT was compared with 3 other functional tests: the Medical Research Council (MRC) muscle test, the 6MWT, and the TUG. Seven days after awakening,[11] manual muscle strength was tested and scored using the MRC muscle test, which is a measure of strength used to quantify muscle weakness; the range of MRC muscle test scores is 0 to 60. A clinical diagnosis of ICU-acquired weakness (ICUAW) is made based upon a sum score of less than 48/60. The reliability of the MRC sum score has been examined in the critically ill population with conflicting results. Hough and colleagues[13] reported poor agreement in the ICU (kappa=.38) and excellent agreement after ICU discharge (kappa=1.0), whereas Hermans and colleagues[14] found good agreement in the ICU (kappa=.68). At ICU discharge, both the 6MWT and the TUG were performed. The 6MWT is a commonly used, simple, and inexpensive submaximal test of physical function. It has been found to correlate moderately with peak oxygen uptake measured by formal exercise testing (r=.5–.7) in cardiorespiratory impaired populations.[15] The TUG is a test of functional mobility in older adults. It measures the time (in seconds) taken to stand from a chair, walk 3 m, and return to the sitting position.[16] For both the 6MWT and TUG, standardized instructions were given, and 2 6MWTs were performed to reduce variability associated with practice effects[17]; the best value was recorded. Participants who used a gait aid during TUG testing also completed the test without a gait aid, and the best value was recorded.

In addition to functional tests, 2 patient-report outcome measures of health-related quality of life (HRQoL)—the 36-Item Short-Form Health Survey version 2 (SF-36v2) and the Assessment of Quality of Life (AQoL)—were included for correlations, as they were used as part of the larger trial at recruitment and at 3, 6, and 12 months post-ICU discharge. Post-hospital discharge outcome measures were performed at hospital outpatient appointments or home visits, as required. The SF-36v2 is an 8-domain, generic health status questionnaire[18] that has been validated[19] and recommended for use[20] in the ICU population. It was administered to participants as close to enrollment in the RCT as possible (day 5 post-ICU admission or later) as a "then test," where participants retrospectively estimated their premorbid HRQoL. The AQoL is a multiattribute utility instrument assessing handicap arising from health states. It consists of 15 questions, each with 4 response levels. The health utility index, the most commonly used item from the AQoL, ranges from 1.00 (best HRQoL state) to −0.04 (worst HRQoL state), where 0.00 is a death-equivalent state.[21] The AQoL has not yet been validated for use in the critical care setting.

Statistical Analyses

The analyses were considered in 2 parts. The first part involved reviewing the components of the original test, assessing the unidimensionality of the PFIT, and developing a score for the test to address the first study aim. The second part involved analyzing the clinimetric properties of the test using the interval score developed in part 1.

Part 1: Development of the PFIT-s. As recommended by Tennant and Pallant,[22] prior to conducting Rasch analysis, exploratory principal component analysis (PCA) was used to broadly assess the dimensionality of the original 5-item PFIT. Principal component analysis was performed on all available admission data (n=144) using varimax rotation. Rasch analysis, based on item response theory, was then used to investigate the unidimensionality of the PFIT. Rasch analysis can provide evidence of construct validity, as it establishes whether a scale is measuring a single unidimensional trait (in this case, function) or is influenced by other constructs (eg, behavior).[23] Response bias (referred to as differential item function) related to the personal attributes (eg, sex, age) of different subgroups of patients can be established, and the difficulty hierarchy of test items in the scale can be evaluated.[24] The RUMM2020 Rasch measurement software program (version 4.0,

1997–2004, RUMM Laboratory Pty Ltd, Duncraig, Australia) was used to perform the Rasch analyses. Overall fit to the model (unidimensionality) was reported if the chi-square value for item trait interaction was greater than .05 and then confirmed using the *t*-test procedure recommended by Tennant and Pallant.[22] Item fit residuals greater than ±2.5 were used to identify multidimensionality or redundancy of items. Differential item functioning was assessed for age (20–49, 50–69, 70+ years), sex, and Acute Physiology and Chronic Health Evaluation II (APACHE II) score (7–17=mild, 18–22=moderate, 23+=severe, as assessed at ICU admission). An ordinal score was derived and conversion to an interval score was achieved using Rasch analysis. Further information explaining Rasch analysis is provided in the eAppendix (available at ptjournal.apta.org).

The ordinal scoring ranges for cadence were developed using clinically sensible cutpoints that approximated the tertile values for each item taken from the complete data set. Due to the importance of antigravity muscle strength in functioning, the ranges for muscle strength were based on the Oxford grading system.[25] Four clinically sensible categories were used for the sit-to-stand item based on the level of assistance required to complete the task. Ceiling and floor effects were calculated as the number of participants who scored the highest or lowest possible score, respectively, divided by the total sample size. No imputation of missing data was undertaken. The newly developed ordinal scale was called the PFIT-s.

Part 2: Clinimetric properties of the PFIT-s. Convergent validity is present when 2 measures believed to reflect a similar underlying construct have a moderate to high correlation.[8] Correlations (*r* and rho) were used to test for convergent validity of the ICU discharge PFIT-s against the TUG and 6MWT, which were measured concurrently at ICU discharge, and the MRC sum score obtained 7 days after awakening.[11] The MRC scores were dichotomized to < (ICUAW) or ≥ (no ICUAW) a score of 48/60.[11] Discriminant validity is present when measures of 2 different constructs present different results, demonstrating the instrument has the ability to discriminate between the constructs.[8] Discriminant validity was measured for the PFIT-s by comparing it with a measure of a different construct: body mass index (BMI). Correlations were defined as: .0 to .25=no relationship, .25 to .5=fair relationship, .5 to .75=moderate to good relationship, and >.75=good to excellent relationship.[8]

Multiple regression analyses were conducted to investigate the predictive utility (the ability of an instrument to predict future health states) of the admission PFIT-s. Linear and binary logistic regression modeling were applied for continuous and dichotomous outcomes, respectively. Potential baseline covariates were patient age, sex, APACHE II score, ventilation status at day 5 post-ICU admission, BMI, SF36v2 physical component summary (PCS) and physical function summary (PFS) scores, and AQoL utility scores. Potential covariates with a significant univariate correlation with outcome and an absence of collinearity were initially included in the regression model and retained if identified as a significant factor in the model. Admission PFIT-s interval score was the variable of interest and was initially included in each regression model. The outcomes of interest in the regression models were: MRC sum score defining ICUAW (<48 or ≥48), discharge destination, acute care hospital length of stay, 28-day or 12-month mortality, ICU or hospital readmission within the 12-month trial follow-up period, AQoL utility score, and SF-36v2 PFS or PCS score (at 3, 6, and 12 months post-ICU discharge).

The effect size index (ESI)[26] was used to calculate measurement responsiveness for the PFIT-s. A positive ESI denotes improvement in health status. Effect size indexes of 0.2, 0.5, and 0.8 have been interpreted to represent small, moderate, and large responsiveness to change, respectively.[27]

The MCID is the minimum change that needs to occur to reflect a clinically meaningful change in patient function. A systematic review performed by Norman et al[28] concluded that half the baseline standard deviation is often a good approximation of the MCID, regardless of whether obtained from distributional or anchor-based methods. This method, therefore, was used to approximate the MCID for the PFIT-s.

Results

The flow of participants through the nested PFIT study and the flow of the PFIT scoring development and analyses are shown in Figures 1 and 2, respectively. Participant demographics are given in Table 2. The median (interquartile range [IQR]) number of days from ICU admission to the performance of the first (admission) PFIT was 6 (5–9), and the median (IQR) number of days between the PFIT measures was 4 (2–10). All available data were used in the analyses. One hundred forty-four participants of the total sample (n=150) completed the admission PFIT, and 116 participants completed the discharge PFIT. The main reasons for missing data at admission were death, confusion, and sedation. No adverse events occurred during any PFIT. Adverse events were defined in our previously published work.[5] At ICU discharge, 119 participants

(79%) completed the 6MWT, and 97 participants (65%) completed the TUG; 27 participants were unable to complete the TUG due to an inability to stand up from the chair.

Part 1: Development of the PFIT-s

The results of PCA of the original 5-item PFIT indicated that all items loaded on the first of 2 components except the shoulder lifts per minute item (see eTab. 1, available at ptjournal.apta.org). Based on the results of the exploratory PCA analysis, Rasch analysis was subsequently conducted both with and without the shoulder lifts item included (Tab. 1).

Fit to the model (unidimensionality). Fit to the Rasch model was improved with: (1) the removal of the shoulder lifts per minute item and (2) removal of disordered thresholds for the shoulder and knee strength items, by rescoring from the 0 to 5 Oxford grading scale to a 0 to 3 score as outlined in Table 3. For the new 4-item PFIT, fit to the model was achieved at ICU admission with a total item chi-square value of 5.89 (df=8, P=.66). Some deviation from the model was identified at discharge (total item χ^2=16.91, df=8, P=.03). However, overall fit to the model was achieved with the pooled dataset (combined admission and discharge data), and unidimensionality was indicated in the admission and discharge datasets using the t-test procedure with point estimates of 3.47% and 1.71%, respectively. There was no DIF (response bias) by age, sex, or APACHE II score at admission. Significant systematic DIF by sex was identified for the maximum knee extension item at discharge (for a given ability, male participants scored systematically higher in knee extension strength than female participants, P<.001).

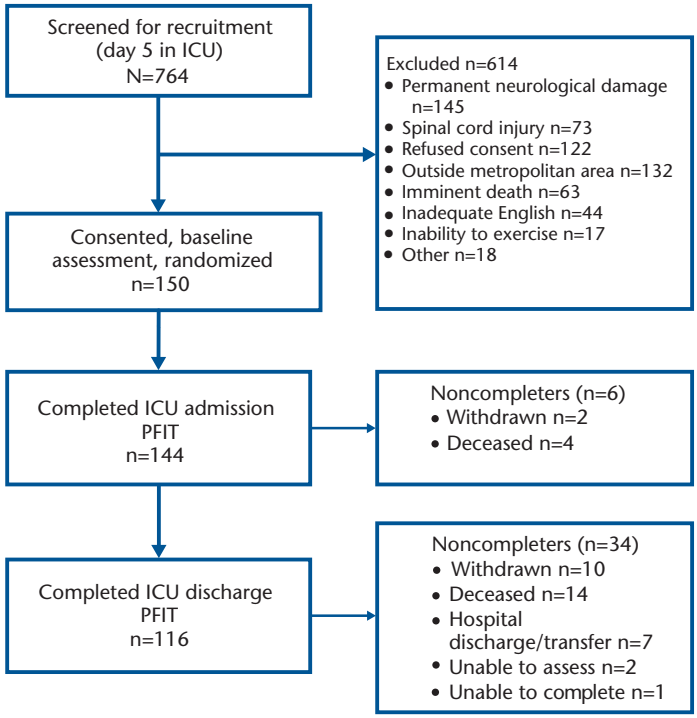

Figure 1.
Participant flow through the nested Physical Function ICU Test (PFIT) study. ICU=intensive care unit. Twenty-eight participants completed ICU admission but not ICU discharge PFIT measures.

Score development. An ordinal scoring system (0–12) and an interval scoring system (0–10) were developed, which were called the PFIT-s to distinguish them from the original PFIT. The ordinal scale was developed based on pooled admission and discharge data from the PFIT-s components using the classification shown in Table 3, where the ordinal score is obtained out of 12 (adding scores out of 3 for the 4 items). The interval score was obtained using Rasch analyses. The ordinal and interval scoring systems and conversion algorithm are given in Table 4. For the PFIT-s, a floor effect of 21.5% was found at admission, as 31 out of a total of 144 participants did not score. A ceiling effect of 22.2% was identified at discharge, as 26 out of 117 participants achieved the highest score.

Part 2: Clinimetric Properties of the PFIT-s

Validity. Convergent validity was present, as a significant moderate correlation was found for the discharge PFIT-s with the TUG (r=−.60, 95% confidence interval [95% CI]=−.70 to −.46, P<.001), the 6MWT (r=.41, 95% CI=0.24 to 0.55, P<.001), and the MRC muscle test (rho=.49, 95% CI=.33 to .62, P<.001, n=105). Low correlation was observed between the BMI and the admission PFIT-s scores (r=−.011, 95% CI=−0.18 to 0.16, n=137), demonstrating divergent validity.

Responsiveness. The ESI for the PFIT-s was 0.82 (95% CI=0.66 to 0.99), which represents a large responsiveness to change.[27] The MCID was calculated to be 0.5 × 3.06 (standard deviation at ICU

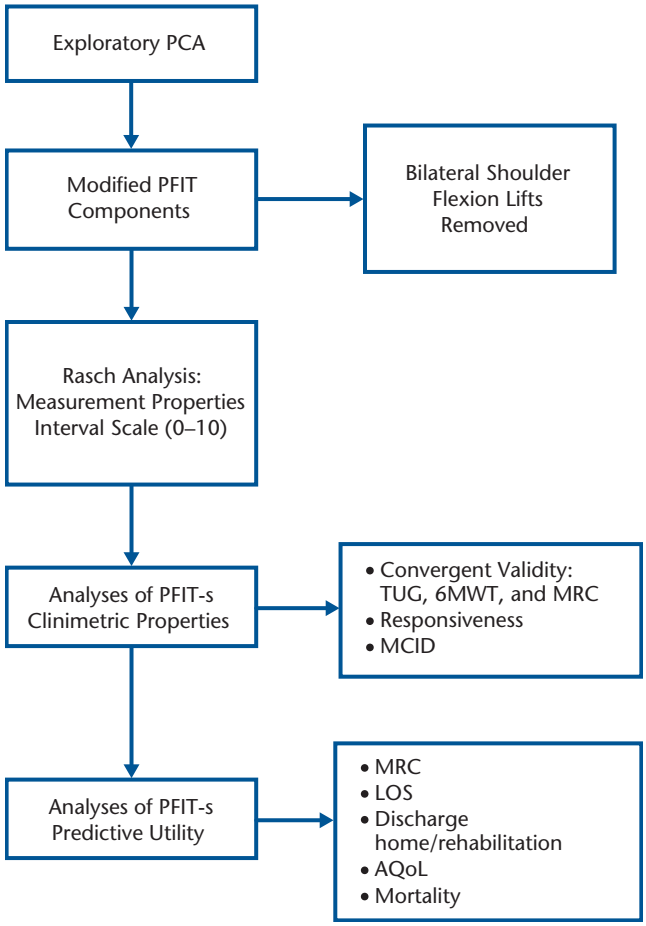

Figure 2.
Flow of Physical Function ICU Test (PFIT) scoring development and analyses. PFIT-s=new 4-item PFIT; PCA=principal component analysis, TUG=Timed "Up & Go" Test (a test of functional mobility in older adults; it records the time taken in seconds for participants to stand from a chair, walk 3 m, return, and sit down again); 6MWT=Six-Minute Walk Test (the test measures the distance that a person can quickly walk on a flat, hard surface in 6 minutes, is self-paced and submaximal, and reflects the functional exercise level for daily physical activities); MRC=Medical Research Council muscle test (a test designed to evaluate muscle strength in which 3 muscle groups of the upper and lower limbs are given a score from 0 [paralysis] to 5 [normal muscle strength]; the range of scores is 0–60, and scores <48 indicate ICU-acquired weakness); MCID=minimal clinically important difference; LOS=length of stay (ICU and acute care hospital); AQoL=Assessment of Quality of Life instrument (a multi-attribute utility instrument comprising 15 items in 5 dimensions assessing health-related quality of life; the AQoL utility score was designed to allow the calculation of quality-adjusted life-years).

admission)=1.5 points on the 10-point interval PFIT-s scale. This value represents 15% of the scale width.

Predictive utility. At ICU admission, the PFIT-s demonstrated predictive utility for several patient and hospital outcomes, and higher PFIT-s scores (better function) were positively associated with: obtaining a higher MRC sum score (being stronger) (≥48) (odds ratio [OR]=1.28, $P<.001$); discharge home (OR=1.20, $P=.01$), and reduced likelihood of discharge to inpatient rehabilitation (OR=0.86, $P=.02$).

Higher admission PFIT-s scores (better function) also were associated with reduced acute care hospital length of stay (B coefficient=−2.13, $P<.001$). Higher admission PFIT-s score and lower age were significant factors in determining AQoL utility scores at the 3-, 6-, and 12-month follow-ups (PFIT-s B coefficient=0.04, $P<.05$) at each time point (see eTab. 2, available at ptjournal.apta.org). The admission PFIT-s did not have predictive ability for ICU or hospital readmission, 28-day or 12-month mortality, or SF-36v2 PCS or PFS score at 3-, 6-, or 12-month follow-up (results not presented). Increasing age (B coefficient=0.03, $P=.03$) and APACHE II score (B coefficient=0.08, $P=.01$) were associated with increased likelihood of 12-month mortality.

Discussion

This study has established that the PFIT-s is a unidimensional, valid, and responsive objective measure of physical function that has moderate correlations with other commonly used functional and strength measures. Previous work by our group established the reliability of the PFIT.[5] The clinimetric testing of the PFIT-s supports its validity and responsiveness, and this is the first

ICU quantitative test of function to be reported that has been compared with other commonly used tests. The PFIT-s can be used to improve clinicians' and researchers' ability to measure the effectiveness of selected treatments and to objectively compare the functional physical capacity of patients across their ICU stay.

The environment of the ICU presents unique challenges to measuring functional outcomes. Measuring muscle strength and function in the ICU has several limitations, as patients need to be awake and cooperative to undertake most of the measures.[29] Success of volitional testing, therefore, will depend on patient characteristics and sedation practices in different units.[30] In Australia, the ratio of nurses to patients, sedation practices, and the role of the physical therapist may differ from other countries,[31] and any research published in an ICU population must consider these differences for the generalizability of the findings.

Several authors have developed instruments to measure function specifically in the ICU setting.[4–6] Zanni et al[6] developed the FSS-ICU using the Functional Independence Measure (FIM) as a guide and chose the most relevant domains for use in the ICU. The FSS-ICU includes 2 tasks from the FIM and 3 other measures. Each task is given a score between 1 (complete assistance) and 7 (complete independence) and assesses: ambulation, rolling, sitting, supine to sitting, and sit-to-stand transfers and was tested in a medical ICU. Although the PFIT-s also includes a form of ambulation (marching in place), the other components are quite different from the FSS-ICU. The original PFIT was developed by our group in 2007, before any other specific tests had been developed and despite little interest from physical therapists in rehabilitation in the ICU or outcomes to measure change in this population. Given that our results demonstrated a floor effect at ICU admission in our population of moderately unwell medical and surgical patients, it is possible that we did not include test items at a low level to cover the ICU population range of abilities. The most common reason for inability to perform the test at this time point was that the patient was not awake. Any volitional test, therefore, would be difficult to perform. Conversely, we also demonstrated a ceiling effect at ICU discharge of similar magnitude, suggesting we need higher-order tasks. These tasks may include walking away from the bed, but further development of the PFIT-s is warranted to address these issues.

Yet, finding one measure of function that is applicable to all patients may not be possible, and use of 2 (or more) different tests may be necessary to measure level of function effectively.[32] The recent publication of use of the FSS-ICU in a long-term acute care hospital did not include full clinimetric testing of this outcome,[7] although the responsiveness

Table 2.
Participant Demographics[a]

Characteristic	Total Sample (n=144)	Clinimetric Sample (n=116)
Age (y), $\overline{X}$ (SD)	60.4 (15.8)	59.3 (15.4)
Sex (% male)	63	60
BMI (kg/m^2), $\overline{X}$ (SD)	27.7 (5.8)	27.7 (5.6)
APACHE II, $\overline{X}$ (SD)	19.3 (6.0)	18.8 (6.0)
ICU diagnosis (%)		
Pneumonia	17.4	15.5
Cardiac	11.8	8.5
Cardiac surgery	22.9	23
Other surgery	15.9	18
Liver disease/transplantation	9.7	11
Cardiac arrest	5.6	5
Sepsis	6.9	6
Renal	3.5	3.5
Other	6.3	9.5
28-d mortality (%)	5.6	0.9
12-mo mortality (%)	19.4	14.7
ICU LOS (d), median (IQR)	7 (6–10)	7 (6–11)
ICU LOS, ≥10 d (%)	33.6	33.6
Acute LOS (d), median (IQR)	22.0 (13.0–36.0)	22.5 (16.0–38.8)
ICUAW, (% yes)	19.4	19.8
MV (h), median (IQR)	98.0 (44.75–169.3)	92.0 (35.5–163.0)
MV at day 5, % (n)	52.8 (76)	49.1 (57)
Readmissions,[b] % (n)	36.1 (52)	50.0 (58)
Discharged to home, % (n)	57.6 (83)	63.8 (74)

[a] BMI=body mass index; ICU=intensive care unit; APACHE II=Acute Physiology and Chronic Health Evaluation, a disease severity scoring system for adults admitted to the ICU; LOS=length of stay; IQR=interquartile range; ICUAW=ICU-acquired weakness; MV=mechanical ventilation.
[b] Readmissions=acute care hospital readmissions during 12-mo study follow-up period.

Table 3.
Classification of Component Scores Used in the Physical Function ICU Test (Scored) (PFIT-s) Ordinal Score

PFIT-s Components			
Assistance	**Cadence (steps/min)**	**Shoulder Strength[a]**	**Knee Strength[b]**
0=unable	0=unable	0=grade 0, 1, or 2	0=grade 0, 1, or 2
1=assist × 2	1=>0–49	1=grade 3	1=grade 3
2=assist × 1	2=50–<80	2=grade 4	2=grade 4
3=no assistance	3=80+	3=grade 5	3=grade 5

[a] Maximum strength of left or right shoulder flexion using the Oxford grading system.
[b] Maximum strength of left or right knee extension using the Oxford grading system.

Table 4.
Ordinal Scores and Equivalent Interval Scores for the Physical Function ICU Test (Scored) (PFIT-s)[a]

Scale	PFIT-s Score												
Ordinal	0	1	2	3	4	5	6	7	8	9	10	11	12
Interval	0	2	3.2	3.9	4.4	4.9	5.4	5.9	6.4	7.1	7.9	8.8	10

[a] Algorithm for conversion from ordinal to interval score=5.418 + (1.068 × logit location of ordinal score).

was assessed to be 0.25 (effect size for entire sample) and the results appear promising. It is early in the development of accurate functional tests in the ICU and beyond ICU discharge, and further research in this area should be and will be forthcoming in the future. A framework for reporting outcomes in ICU related to the World Health Organization's *International Classification of Functioning, Disability and Health* (ICF) model has been suggested.[33] Using this model will allow clinicians to choose one or more tests to assess the activity limitations of patients using, for example, the PFIT-s or 6MWT. These test results can inform rehabilitation and post–acute care needs.

In order to perform the original test, patients needed to be out of bed in a sitting position. With the PFIT-s, it is possible to perform isolated tasks and still obtain a score. For example, if a patient cannot move out of bed, strength testing can be performed with the patient in bed or sitting on the edge of the bed. The lower score obtained is reflective of the acuity of the patient at the time of measurement. Additionally, components of the test can be performed earlier if the patient is assessed while awake. It is not necessary to wait until 7 days after awakening as we did in this study. This approach was followed to obtain PFIT measurements at the same time, as recommended for MRC muscle test measurements.[11]

Providing an interval score using RASCH analyses allows a more precise and sensitive measure of change compared with an ordinal score within and across individuals.[8,34] It also provides a method for more accurately measuring and monitoring clinical patient changes than ordinal measures. However, in a clinical situation with no means of converting to an interval score, the ordinal score can easily be used by clinicians at the bedside.

Unidimensionality is an important attribute of a measure. Functional scales should reflect one construct, making the comparison between scores in individuals more valid.[8] The PFIT-s measured only function, as demonstrated using the combination of admission and discharge data. The PFIT-s scores correlated moderately with the MRC muscle test scores, perhaps because there are 2 measures of strength within the PFIT-s. Conversely, there may be a true correlation between strength and function in this population. In comparing the PFIT-s with the MRC, the fact that there is controversy regarding the reliability and utility of within-ICU measurements obtained with the MRC muscle test must be taken into account,[13,14,35] as should the fact that currently both isometric and through-range techniques are utilized to test strength. Variability between raters on the MRC muscle test may alter the correlation of this measure with other tests.

The PFIT-s also demonstrated predictive ability and may facilitate the identification of patients who are more likely to require rehabilitation and those who are more likely to have improved HRQoL (as measured using the AQoL) after discharge. This finding presents the possibility of targeting of scarce resources both in the ICU and beyond. If these patients are targeted with intensive physical therapy early in the ICU, outcomes

may be improved later, reducing time in ongoing rehabilitation.[35,36] Use of the PFIT-s to predict patient outcomes warrants further research.

The added advantage of the PFIT-s, in combination with use of the Borg Scale of Perceived Exertion, is that it allows the objective prescription of exercise at an appropriate level for the patient to achieve a training effect. For example, a percentage of the time of marching in place in the test can be used to commence subsequent rehabilitation sessions, and this time can be increased as the patient improves. This successful method of exercise prescription is used in other patient populations such as in pulmonary rehabilitation, where the 6MWT is used to prescribe exercise as well as evaluate outcome.[17,37] Prescribing exercise in this way assists in training patients at an adequate level for their ability and to achieve a training response at a given point in time. Using this approach to exercise prescription intensity was safe in an Australian ICU setting, with no serious adverse events recorded.[38]

Limitations

There were missing data for the MRC muscle test, as it was difficult to perform this test at ICU admission due to patients being sedated. This difficulty was reported in a recent study.[29] We acknowledge that the measure of MCID using the distribution-based method of Norman and colleagues[28] in this study may be criticized.[39,40] However, there is currently no consensus for defining the MCID, although it is common to use several different methods.[40] It is argued that MCID values are designed to determine the clinical significance of changes in individual participants, and translating these values to group mean scores (or extrapolating even further to between-group differences) may not be valid.[39,41] The MCID is also applied across a given score range. However, the score range may vary with disease severity, and this limitation may need to be considered.[40] Another limitation is that, like the FIM, the scoring of the assistance level provided to the patient in the sit-to-stand task is somewhat subjective despite standardized methods and instructions. The amount of assistance is scored based on ability of the patient to stand and the 2 therapists' subjective assessment of the amount of help required. We determined the cadence cutpoints for the ordinal scale based on tertiles in our population. These cutpoints may not be generalizable to other populations with varying levels of illness, and further research should be undertaken in different populations of patients.

Conclusions

The PFIT-s measured between days 5 to 10 of ICU admission in a sample of participants who were in the ICU for a minimum of 5 days was simple, inexpensive, and had high clinical utility. It was shown to measure one construct, to have validity compared with commonly used functional tests, and to be predictive of several important patient parameters related to function. However, floor and ceiling effects existed. We recommend, therefore, that the PFIT-s be adopted for use to measure physical function in patients in the ICU or in those discharged to longer-term care in combination with other measures. Future research should be aimed at identifying several tests that may define physical function and could be used as a test battery to measure activity limitations, taking into consideration floor and ceiling effects of each test for survivors of the ICU.

Dr Denehy, Dr Skinner, and Dr Berney provided concept/idea/research design. Dr Denehy, Dr de Morton, Ms Edbrooke, and Dr Berney provided writing. Dr Skinner and Ms Haines provided data collection. Dr Denehy, Dr de Morton, and Ms Edbrooke provided data analysis. Dr Denehy, Ms Edbrooke, and Dr Berney provided project management. Dr Denehy and Dr Berney provided fund procurement. Ms Haines and Dr Warrillow provided study participants. Ms Haines and Dr Warrillow and Dr Berney provided institutional liaisons. All authors provided consultation (including review of manuscript before submission). The authors thank Austin Health Melbourne for providing facilities/equipment.

Approval for the RCT was obtained from the Human Research Ethics Committee of Austin Health, Melbourne, Australia.

The RCT is registered with the Australian and New Zealand Clinical Trials Registry (ACTRN 12605000776606).

A poster presentation of this work was given at the American Thoracic Society International Conference; May 13–18, 2011; Denver, Colorado.

DOI: 10.2522/ptj.20120310

References

1 Burtin C, Clerckx B, Robbeets C, et al. Early exercise in critically ill patients enhances short-term functional recovery. *Crit Care Med.* 2009;37:2499–2505.

2 Morris PE, Goad A, Thompson C, et al. Early intensive care unit mobility therapy in the treatment of acute respiratory failure. *Crit Care Med.* 2008;36:2238–2243.

3 Schweickert WD, Pohlman MC, Pohlman AS, et al. Early physical and occupational therapy in mechanically ventilated, critically ill patients: a randomised controlled trial. *Lancet.* 2009;373:1874–1882.

4 DiCicco J, Whalen D. University of Rochester Acute Care Evaluation: development of a new functional outcome measure for the acute care setting. *J Acute Care Phys Ther.* 2010;1:14–20.

5 Skinner EH, Berney S, Warrillow S, Denehy L. Development of a physical function outcome measure (PFIT) and a pilot exercise training protocol for use in intensive care. *Crit Care Resusc.* 2009;11:110–115.

6 Zanni JM, Korupolu R, Fan E, et al. Rehabilitation therapy and outcomes in acute respiratory failure: an observational pilot project. *J Crit Care.* 2009;25:254–262.

7 Thrush A, Rozek M, Dekerlegand JL. The clinical utility of the Functional Status Score for the Intensive Care Unit (FSS-ICU) at a long-term acute care hospital: a prospective cohort study. *Phys Ther.* 2012;92:1536–1545.

8 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. East Norwalk, CT: Appleton & Lange; 2009.

9 Denehy L, Berney S, Skinner E, et al. Evaluation of exercise rehabilitation for survivors of intensive care: protocol for a single blind randomised controlled trial. *Open Crit Care Med J.* 2008;1:39–47.

10 Linacre J. Sample size and item calibration stability. *Rasch Measurement Transactions.* 1994;7:328.

11 De Jonghe B, Sharshar T, Lefaucheur J, et al. Paresis acquired in the intensive care unit: a prospective multicenter study. *JAMA.* 2002;288:2859–2867.

12 Borg G. Ratings of perceived exertion and heart rates during short-term cycle exercise and their use in a new cycling strength test. *Int J Sports Med.* 1982;3:153–158.

13 Hough CL, Lieu BK, Caldwell ES. Manual muscle strength testing of critically ill patients: feasibility and interobserver agreement. *Crit Care.* 2011;15:R43.

14 Hermans G, Clerckx B, Vanhullebusch T, et al. Interobserver agreement of Medical Research Council sum-score and handgrip strength in the intensive care unit. *Muscle Nerve.* 2012;45:18–25.

15 Cahalin L, Pappagianopoulos P, Prevost S, et al. The relationship of the 6-min walk test to maximal oxygen consumption in transplant candidates with end-stage lung disease. *Chest.* 1995;108:452–459.

16 Podsiadlo D, Richardson S. The timed "Up and Go" test: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* 1991;39:142–148.

17 Jenkins S. 6-Minute Walk Test in patients with COPD: clinical applications in pulmonary rehabilitation. *Physiotherapy.* 2007;93:175–182.

18 Ware J. *SF-36 Health Survey Manual and Interpretation Guide*. Boston, MA: The Medical Outcomes Trust; 1993.

19 Chrispin PS, Scotton H, Rogers J, et al. Short Form 36 in the intensive care unit: assessment of acceptability, reliability and validity of the questionnaire. *Anaesthesia.* 1997;52:15–23.

20 Black NA, Jenkinson C, Hayes JA, et al. Review of outcome measures used in adult critical care. *Crit Care Med.* 2001;29:2119–2124.

21 Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life. *Qual Life Res.* 1999;8:209–224.

22 Tennant A, Pallant J. Unidimensionality matters! (A tale of two Smiths?). *Rasch Measurement Transactions.* 2006;20:1048–1051.

23 Bond T, Fox C. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates; 2001.

24 Miller KJ, Slade AL, Pallant JF, Galea MP. Evaluation of the psychometric properties of the upper limb subscales of the Motor Assessment Scale using a Rasch analysis model. *J Rehabil Med.* 2010;42:315–322.

25 Daniels L, Worthingham C. *Muscle Testing: Techniques of Manual Examination*. 5th ed. Philadelphia, PA: FA Davis Co; 1995.

26 Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:S178–S189.

27 Husted J, Cook R, Farewell V, Gladman D. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;53:459–468.

28 Norman G, Sloan J, Wyrwich K. Interpretation of changes on health related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41:582–592.

29 Connolly B, Denehy L, Brett S, et al. Exercise rehabilitation following hospital discharge in survivors of critical illness: an integrative review. *Crit Care.* 2012;16:226.

30 Hough CL, Needham DM. The role of future longitudinal studies in ICU survivors: understanding determinants and pathophysiology of weakness and neuromuscular dysfunction. *Curr Opin Crit Care.* 2007;13:489–496.

31 Berney S, Haines K, Skinner EH, Denehy L. Safety and feasibility of an exercise prescription approach to rehabilitation across the continuum of care for survivors of critical illness. *Phys Ther.* 2012;92:1524–1535.

32 Denehy L, Elliott D. Strategies for post ICU rehabilitation. *Curr Opin Crit Care.* 2012;18:503–508.

33 Iwashyna TJ, Netzer G. The burdens of survivorship: an approach to thinking about long-term outcomes after critical illness. *Semin Respir Crit Care Med.* 2012;33:327–338.

34 Stevens RD, Marshall SA, Cornblath DR, et al. A framework for diagnosing and classifying intensive care unit-acquired weakness. *Crit Care Med.* 2009;37(10 suppl):S299–S308.

35 Fan E. What is stopping us from early mobility in the intensive care unit? *Crit Care Med.* 2010;38:2254–2255.

36 Herridge M. Long-term outcomes after critical illness: past, present, future. *Curr Opin Crit Care.* 2007;13:407–423.

37 American Thoracic Society. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med.* 2002;166:111–117.

38 Berney S, Haines K, Skinner E, Denehy L. An exercise prescription approach to rehabilitation for survivors of critical illness. *Phys Ther.* 2012;92:1524–1535.

39 Ferreira ML, Herbert RD. Clinical importance of an intervention must reside with the patient. *Aust J Physiother.* 2009;55:219.

40 Swartz RJ, Schwartz C, Basch E, et al. The king's foot of patient-reported outcomes: current practices and new developments for the measurement of change. *Qual Life Res.* 2010;20:1159–1167.

41 Dolmage T, Hill K, Evans R, Goldstein R. Has my patient responded? Interpreting clinical measurements such as the 6-minute-walk-test. *Am J Respir Crit Care Med.* 2011;184:642–646.

eAppendix.

Rasch Analysis Information[a]

Rasch analysis is a measurement model that is used to assess the psychometric properties of a rating scale, in particular unidimensionality.

Data from individual items of a scale, which are summed to give a total score, are compared with expected measures using the Rasch model. A number of fit statistics are used to test how well the observed scores fit the Rasch model's expected scores. For example, with a dichotomous questionnaire item used to measure anxiety, the probability of a person responding positively to an item is a function of the difference between his or her anxiety level and the anxiety level expressed by a positive response to the question.

Item-person interaction statistics are transformed to z scores; if the items and patients fit the model, z scores approximate the normal distribution (mean=0, standard deviation=1). The chi-square statistic also is used to assess the hierarchical ordering of questionnaire items (item-trait interaction).

Causes of deviations from expected scores are investigated. For example, in a well-fitting model, individuals with higher levels of anxiety should respond with higher scores, and those with lower anxiety levels should respond with lower scores. Among other reasons, poor model fit may be a result of too many item response options or wording of items causing the patient's item responses to be inconsistent.

The Rasch model can be used to test for:

- unidimensionality of items—a requirement of items in a scale. This property can be assessed by looking at residual patterns (differences between expected and observed scores) and differential item functioning (DIF). Residuals between ± 2.5 indicate acceptable fit to the model.
- category ordering—do item response options in a particular question behave how they should?
- DIF (item bias)—the probability of a person responding positively to an item should not be dependent on a particular trait (eg, sex, body mass index). For example, male and female patients with the same anxiety levels should demonstrate the same probability of responding positively to an anxiety level item. If the probability between sexes is not the same, the questionnaire item can be said to display DIF by sex, which violates the unidimensionality requirement.

A linear transformation (data on an interval scale) can be obtained from the raw ordinal data if fit to the Rasch model is achieved, allowing parametric statistical analyses to be performed on data that are normally distributed.

[a] Pallant J, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depresssion Scale (HADS). *Br J Clin Psychol.* 2007;46:1–18.

eTable 1.

Principal Components Analysis (Component Matrix[a])

	Component	
	1	2
Assistance	.847	.399
Cadence	.891	.308
ShoulderMax	.512	.807
KneeMax	.629	.699
Lifts		.906

[a] Varimax rotation with Kaiser normalization. Shoulder lift item was the only item that did not load on the first component.

eTable 2.

Predictive Equations for Acute Care Hospital Length of Stay and Assessment of Quality of Life (AQoL) Utility Scores

Variable	Predictive Equation
Acute care hospital length of stay	41.292 – (admission PFIT[a] score × 2.126)
3-mo AQoL utility score	0.819 + (0.036 × PFIT score) – (0.008 × age)
6-mo AQoL utility score	0.806 + (0.035 × PFIT score) – (0.007 × age)
12-mo AQoL utility score	0.827 + (0.027 × PFIT score) – (0.007 × age)

[a] PFIT=Physical Function ICU Test.