Comprehensive Physical Function Measure for the Intensive Care Unit: The use of Rasch Analysis and Item Response Theory

Michelle Lynn Peterson
Nova Southeastern University

Follow this and additional works at: https://nsuworks.nova.edu/hpd_pt_stuetd

Part of the Physical Therapy Commons

All rights reserved. This publication is intended for use solely by faculty, students, and staff of Nova Southeastern University. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, now known or later developed, including but not limited to photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author or the publisher.

NSUWorks Citation

This Dissertation is brought to you by the Department of Physical Therapy at NSUWorks. It has been accepted for inclusion in Department of Physical Therapy Student Theses, Dissertations and Capstones by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.
A Comprehensive Physical Function Measure for the Intensive Care Unit: The use of Rasch Analysis and Item Response Theory

By:
Michelle Lynn Peterson

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Nova Southeastern University
Dr. Pallavi Patel College of Health Care Sciences
Department of Physical Therapy

October 2020
We hereby certify that this dissertation, submitted by Michelle Lynn Peterson, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirement for the degree of Doctor of Philosophy in Physical Therapy.

______________________________  ______________________
Dr. Bini Litwin PT, DPT, PhD, MBA  Date
Chairperson of Dissertation Committee

______________________________  ______________________
M. Samuel Cheng PT, MS, ScD  Date
Member, Dissertation Committee

______________________________  ______________________
George Fulk PT, PhD  Date
Member, Dissertation Committee

Approved:

______________________________  ______________________
M. Samuel Cheng PT, MS, ScD  Date
Director, Physical Therapy Ph.D Program

______________________________  ______________________
Shari Rone-Adams PT, MHSA, DBA  Date
Chair, Department of Physical Therapy

______________________________  ______________________
Stanley H Wilson PT, EdD, CEAS, FASAHP  Date
Dean and Associate Professor
Abstract

Introduction

This dissertation study provides an in-depth examination of current intensive care unit (ICU) physical function assessment measures and synthesizes these into one comprehensive measurement tool that addresses multiple areas of function.

Background

A recent systematic review identified 26 measures being used in ICU clinical research, although a subsequent systematic review revealed only 14 of those physical function assessment measures have psychometric properties evaluated specifically for the ICU setting. A robust physical function assessment measure for use in the ICU, allows for determination of efficacy of treatment, comparison of studies across settings, and broader interpretation of results.

Specific Aims

Aim 1: Identify physical-function measures currently utilized in the ICU that have been psychometrically tested.

Aim 2: Analyze all measure constructs to determine redundancies and appropriateness for use in the ICU setting according to Rasch analysis and item response theory.

Aim 3: Create a comprehensive, robust functional measurement tool for use with patients in the intensive care unit.
Methodology

Rasch analysis was used for individual activity task evaluation, ranking of task difficulty, and removal of duplicate tasks. IRT Rasch analysis included: item fit, hierarchy, reliability, dimensionality, DIF, and probability. Receiver operating characteristics curve was conducted for predictive validity.

Results

Fifteen items out of a total of 53 met the requirements for an optimal rating scale. The items were ranked according to difficulty and there was no misfit. The reliability indexes were 5.13, $\alpha = .96$ and 21.52, $\alpha = 1.00$ for person and item scores respectively confirming scale hierarchy. The Cronbach Alpha (KR-20) person raw score "test" reliability was 0.96 with SEM 2.72. DIF was deemed non-significant and the probability curves were well delineated and ordered. The comprehensive physical function measure was found to have predictive validity for discharge to home with an optimal ICU admission cut-off score of 42 raw & 51 equal-interval, (sensitivity 71.7%, specificity 67.7%), and an optimal ICU discharge cut-off score of 54 raw & 61 equal-interval, (sensitivity 81.6%, specificity 82.0%).

Summary

A robust, reliable, and valid 15-item comprehensive physical function measure for use in the ICU was developed through Rasch analysis and item response theory.
Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my dissertation Chair, Dr. Bini Litwin. Her continual support included patience, motivation, enthusiasm, and immense knowledge. Her guidance has been invaluable.

Besides my Chair, I would like to thank the rest of my committee, Drs. Samuel Cheng, and George Fulk. Both have offered me encouragement, insightful comments, and the immense support I needed with the difficult methodology that I chose to tackle.

I would be remiss in not also extending my unending gratitude to my colleagues Stacey Jarrell, Laurie Funk, and Gina Ragonese. Stacey, as my Director, always pushed me to succeed and supported my endeavors of research despite the ever-present push of productivity and limited time. I also extend a thank you to Gina for running from floor to floor to consent patients for Laurie and me. I know we drove you crazy at times. Laurie, I cannot begin to say enough “Thank You’s.” You stepped out of your comfort zone to help me with research, being left with a lot of nightly homework. I will never be able to demonstrate my appreciation enough.

Last, but certainly not least, I have to thank my family. My parents, Ed and Monica Peterson, my Grandparents, Fred, and Irmgard Bardenhagen, for their never-failing love and support. And, “Yes, Dad, I’m finally done.” Also, my friends, Michelle Edling and Mary Cook who had to endure all these years of my complaining and limited time. You are all so incredibly special to me.
Table of Contents

Abstract ........................................................................................................................................ iii
Acknowledgement ...................................................................................................................... v
Table of Contents ....................................................................................................................... vi
List of Tables .............................................................................................................................. xi
List of Figures ............................................................................................................................ xii
Chapter 1: Statement of the Problem ......................................................................................... 1
   Introduction ............................................................................................................................ 1
   Background ........................................................................................................................... 1
   Early Mobility in the ICU ...................................................................................................... 4
   Physical Therapy in the ICU ................................................................................................. 5
   Physical Function Outcome Measures in the ICU ............................................................... 7
   Statement of the Problem ...................................................................................................... 10
   Relevance and Significance of the Study .............................................................................. 12
   Specific Aims ......................................................................................................................... 15
   Summary ............................................................................................................................... 16
   Definitions ............................................................................................................................ 17
Chapter 2: Review of Literature ................................................................................................. 20
   Introduction ............................................................................................................................ 20
   Historical Overview ........................................................................................................... 20
   Physical Therapy in the ICU ................................................................................................. 24
      Barriers to Physical Therapy in the ICU ........................................................................... 34
   Outcome Measures for Physical Therapy ............................................................................ 35
   Use of Physical Function Outcome Measures in the ICU ................................................... 39
   ICU Physical Function Outcome Measures ......................................................................... 43
      Physical Function in Intensive Care Unit Test .................................................................. 43
      Chelsea Critical Care Physical Assessment Tool ........................................................... 45
Perme ICU Mobility Score ........................................................................................................ 46
Surgical Intensive Care Unit Optimal Mobility Score ....................................................... 47
ICU Mobility Scale ............................................................................................................. 48
Functional Status Score for the ICU .................................................................................. 49
Acute Care Index of Function ............................................................................................. 50
de Morton Mobility Index .................................................................................................... 51
Short Physical Performance Battery .................................................................................. 52
Early Functional Abilities ..................................................................................................... 53
Functional Capacity Scale .................................................................................................. 54
Medical Research Council Sum Score/Hand Grip ............................................................. 55
Functional Assessment of Burns ......................................................................................... 56
What is Known and Not Known .......................................................................................... 57
Assessment of Medical Acuity ............................................................................................. 60
Rehabilitation/Mobilization Criteria .................................................................................... 60
The Richmond Agitation-Sedation Scale .......................................................................... 61
Acute Physiology and Chronic Health Assessment Evaluation .......................................... 63
Sequential Organ Failure Assessment ................................................................................. 63
Statistical Analysis Review ................................................................................................. 64
Classical Test Theory .......................................................................................................... 64
Item Response Theory ......................................................................................................... 65
Rasch Analysis ...................................................................................................................... 65
Statistical Analysis for this Dissertation ............................................................................. 66
Summary .............................................................................................................................. 67
The Contributions of this Study ........................................................................................... 68

Chapter 3: Methodology ....................................................................................................... 69
Introduction to the Methodology Chapter ........................................................................ 69
Research Method ................................................................................................................ 69
Specific Procedures Employed ............................................................................................ 69
Reliability ........................................................................................................... 97
Scale Dimensionality ....................................................................................... 97
Differential Item Functioning ........................................................................... 99
Item Deduction Process ................................................................................... 100
Rasch Analysis for the 15-Item CPFM ............................................................... 103
  Item Fit ........................................................................................................... 103
  Hierarchy ....................................................................................................... 105
  Reliability ..................................................................................................... 106
  Scale Dimensionality .................................................................................... 106
  Differential Item Functioning ....................................................................... 108
  Probability Curves ....................................................................................... 109
  Scoring .......................................................................................................... 112
Predictive Validity of the 15-Item CPFM .......................................................... 115
Summary of Results .......................................................................................... 117

Chapter 5: Discussion ....................................................................................... 118
  Introduction to the Chapter .......................................................................... 118
  Aim 1 ............................................................................................................ 118
  Aim 2 ............................................................................................................ 120
  Aim 3 ............................................................................................................ 127
  Relevance to Clinical Practice ....................................................................... 142
  Recommendations for Future Research ....................................................... 150
  Limitations and Delimitations ....................................................................... 151
    Limitations .................................................................................................. 151
    Delimitations ............................................................................................... 152
  Summary ........................................................................................................ 153

Appendices ....................................................................................................... 154
  Appendix A: ICU Specific Article Characteristics ......................................... 154
  Appendix B: Cosmin Scores .......................................................................... 158
List of Tables

Table 1: Physical Function Measure Components .......................................................... 57
Table 2: Richmond Agitation-Sedation Scale .................................................................. 62
Table 3: The 53-Item CPFM ................................................................................................. 76
Table 4: Inter-Rater Patient Demographic and Clinical Data ......................................... 89
Table 5: Inter-Rater Reliability Scores ............................................................................... 90
Table 6: Study Patient Demographic and Clinical Data .................................................. 92
Table 7: Hand Grip Likert Scale Values ........................................................................... 94
Table 8: Differential Item Functioning: (53-Item) Gender .............................................. 99
Table 9: Differential Item Functioning: (53-Item) Diagnosis ......................................... 99
Table 10: Outfit Values of the 15-Item CPFM ................................................................. 105
Table 11: Differential Item Functioning (15-Item): Gender ........................................ 108
Table 12: Differential Item Functioning (15-Item): Diagnosis ....................................... 108
Table 13: Observed Counts for Item Responses ............................................................ 110
Table 14: Observed Score and Andrich Values for Bridging and Pen from Floor .......... 111
Table 15: Scoring of the 15-Item CPFM ......................................................................... 112
Table 16: The 15-Item CPFM ............................................................................................. 113
Table 17: Predictive Validity of the 15-Item CPFM ........................................................ 115
Table 18: CPFM versus CPAx: Respiratory Function and Coughing Likert Scale ........ 125
Table 19: Predictive Validity Comparison – Logistic Regression .................................. 136
Table 20: Predictive Validity Comparison - ROC Analysis ............................................ 139
Table 21: Comparison of Patient Characteristics .............................................................. 140
Table 22: ICU Outcome Measure Floor and Ceiling Effects ........................................ 147
Table 23: Physical Function Measure Component Comparison .................................... 149
List of Figures

Figure 1: Participant Recruitment ................................................................. 91
Figure 2: Wright Map of the 53-Item CPFM ................................................... 95
Figure 3: 53-Item CPFM Scale Dimensionality .............................................. 98
Figure 4: Process Summary ............................................................................ 102
Figure 5: Wright Map of the CPFM-15 .......................................................... 104
Figure 6: 15-Item CPFM Scale Dimensionality .............................................. 107
Figure 7: Probability Curve of the CPFM-15 .................................................. 110
Figure 8: ROC Curve Admission Data ........................................................... 116
Figure 9: ROC Curve Discharge Data ............................................................. 116
Figure 10: Wright Map Maximum Probability for the CPFM-15 ..................... 129
Figure 11: Likelihood Ratio Nomogram on Admission to the ICU ................. 133
Figure 12: Likelihood Ratio Nomogram on Discharge from the ICU .............. 134
Chapter 1: Statement of the Problem

Introduction

This dissertation research study provides an in-depth examination of current intensive care unit (ICU) physical function assessment measures for the purpose of synthesizing them into one comprehensive measurement tool that addresses multiple areas of function. Through Rasch analysis and item response theory (IRT), all individual tasks within the physical function measures were analyzed and thoroughly reviewed for inclusion into a final physical function measurement tool. This physical function assessment measure can provide a robust and reliable means to gauge effectiveness of care and to monitor outcomes for patients in the ICU.

Background

As healthcare delivery has advanced and costs have soared from a systemic perspective over the past decades, care of the critically ill patient has shown associated advances in both quality of care and costs. The need to be accountable for those costs has become the focus of current healthcare policy and practice. Arnold Relman reports that we are now in the third medical care system revolution, directed towards understanding current efficiency and efficacy of our healthcare system.\(^1\) The first revolution, beginning in the late 1940’s, was considered the Era of Expansion with growing physician numbers, hospital facilities, insurance coverage, and new developments within science and technology.\(^1\) The second was the Era of Cost Containment to combat inflated costs of healthcare.\(^1\) From 1980-2011, the United States was ranked last of 11 nations on efficiency, equity, and outcomes.\(^2\) Healthcare had gone from 4% of the gross national product to 11% totaling health expenditures per capita of $8,508.00.\(^1,2\) The third revolution is now the Era of Assessment and Accountability associated with an increased need to know more about the costs, safety, and effectiveness of medical treatments. “We can no
longer afford to provide healthcare without knowing more about its successes and failures.”¹ The drive is for an equitable healthcare system, satisfactory quality, and an affordable price. Outcome measures are essential to provide the accountability needed to ensure that a quality driven, affordable, and equitable healthcare system is available from both a systemic and individual patient perspective.

Nowhere is the need greater than in the ICU where the focus of care is to restore physiologic stability and prevent death,³ which requires a multi-faceted approach and a high level of accountability. According to the Centers for Medicare and Medicaid services (CMS), in 2016 the national health expenditure grew another 4.3% to $3.3 trillion dollars, or $10,348.00 per person, and accounted for 17.9% of the gross domestic product.⁴ Patients admitted to the ICU account for approximately 20% of all acute care admissions.⁵ According to the American Hospital Association’s 2014 annual survey, 55,000 critically ill patients are cared for on a daily basis,⁵ accounting for 13.4% of all U.S. hospital costs.⁶ Between the years of 2000-2005, annual critical care medical costs increased by 44.2% (from $56.6 to $81.7 billion).⁷ The Society of Critical Care medicine, in its 2018 critical care updates, reported an overall mortality rate for adult ICU patients ranging from 10% to 29%.⁵ Wunsch and colleagues⁸ indicated a higher mortality risk among those discharged after an ICU stay than those from other hospital units. This was due to both their illness as well as their likelihood for prolonged bed rest.

An ICU stay can be viewed as a continuum that commonly commences with acute clinical deterioration, treatment and care in the ICU, discharge from the ICU, and, ultimately, discharge from the hospital.⁹ Delays in physical recovery may result from severity of the current illness, treatments administered, such as medications, ventilation, hemodialysis; and/or the
secondary complications of bed rest. Failure to identify or address potentially preventable threats to patients anywhere along the continuum can delay recovery.

Prolonged bed rest and immobilization is the number one preventable threat that can lead to loss of strength secondary to muscle disuse, loss of bone density secondary to decreased weight bearing, and increased risk of venous thromboembolism secondary to decreased muscular contraction and mobility. Pressure ulcers can form because of the disruption of blood supply to the tissues, prolonged pressure, and poor mobilization. Pulmonary detriments include increased stasis of secretions and decreased respiratory excursion, leading to pneumonias or atelectasis. From a cardiovascular standpoint, there is decreased plasma volume, orthostatic instability, and decreased venous flow back to the heart. Lack of activity can reduce stroke volume with submaximal exercise up to 30%. From a psychological standpoint, there are occurrences of delirium, depression, and even post-traumatic stress disorder associated with an ICU stay.

Recent investigations have led to identification of ICU-acquired weakness (ICUAW), known as critical illness myopathy or polyneuropathy. ICUAW is characterized as profound weakness greater than what would be expected for usual bed rest. Incidence of ICUAW varies substantially and has been reported ranging from 25% to 100% of patients depending on the severity of the underlying disease. Penuelas et al report a 3% incidence, but highlight the significant association with failure to wean from a ventilator and increased ICU mortality.

Complications experienced by survivors of an ICU stay, classified as Post Intensive Care Syndrome (PICS), are associated with deterioration of strength, physical capabilities, and psychological abilities. The patient with PICS may have a reduced capacity to function, such as returning to work, ambulating community distances, or performing ADLs, decreasing overall
quality of life.\textsuperscript{20} PICS is identified not only immediately post discharge from the hospital, but longitudinal studies show continued effects 1-5 years following hospital discharge.\textsuperscript{20-22} Routine assessment of strength and mobility deficits in the acute care setting can identify patients at risk for PICS.

Early patient mobility is one part of the multi-faceted approach that focuses on reducing the length of stay, improving recovery, promoting functional return, and improving quality of life at discharge.\textsuperscript{5,9,23,24} Early patient mobility also limits neuromuscular dysfunction associated with prolonged bed rest, which leads to significant debility in both physical and psychological domains.\textsuperscript{24} The need for skilled care upon discharge, and the inability to be discharged home due to functional reasons or the lack of care available through family or friends has been shown to be a strong predictor of mortality risk.\textsuperscript{8} Given the negative consequences associated with immobility, rapid and accurate identification of impaired function, early mobility, and discharge planning with physical therapy (PT), are critical to achieving the maximum recovery of the critically ill patient, making rehabilitation a focused strategy.\textsuperscript{9,25,26}

**Early Mobility in the ICU**

Despite the considerable attention in clinical and scientific literature over the last several years and understanding the detrimental effects of bed rest, prolonged immobility still occurs in many ICUs.\textsuperscript{3,25,27} Bed rest, vital organ dysfunction (such as sepsis or hypoxemia), long-term mechanical ventilation, and even sedative medications contribute largely to muscle weakness, functional impairments, and loss of quality of life.\textsuperscript{3,28} The debilitating effects of bed rest and critical illness itself can be mitigated by allowing patients to engage in activity that stimulates both the mind and body.\textsuperscript{25} Early mobility followed by early introduction of physical and occupational therapy shows improvement in functional independence when compared with
control groups. Benefits of early mobility are also noted in increased ambulation distance, ability to perform activities of daily living, and improvement in respiratory function reducing mechanical ventilation time.3,22

The research for outcomes based on mobility, however, is very limited in randomized controlled trials, which limits the strength of current evidence.3,29 Conclusions can be made that early mobility is safe and feasible,3,29,30 but more specifics parameters such as intensity, frequency, and dose of interventions cannot be accurately reported at this time due mostly to the variability of outcome measures utilized within each specific study.3,17,22,25 The lack of consistency in use of an outcome measure limits generalizability and comparison across settings.17 There appears to be a clear need for a standardized outcome measure to objectively capture a patient’s functional status, track effectiveness of treatment interventions, and allow for comparison of outcomes among patients in different ICU settings.25

**Physical Therapy in the ICU**

Physical therapists have not been consistently present in the ICU setting as patients are often considered “too critically ill” to undergo physical rehabilitation. By taking the patient out of bed, mobilization imparts a gravitational force as well as an exercise stimulus, which elicits an acute physiologic response.31 Circulation, perfusion, ventilation, muscle metabolism and alertness are all augmented.31 Reported adverse events during mobilization can include hemodynamic instability, falls, removal of lines or catheters, dislodging of life support, and even death.13,25,29-31

The role of the physical therapist is to “identify and differentiate underlying health conditions, body and system impairments, contextual factors, activity limitations, and
participation restrictions to address the impact on the patient’s function. Physical therapists are uniquely qualified to appropriately prioritize and implement interventions to maximize performance. The ICU physical therapist traditionally focuses on physical impairments and makes recommendations for positioning, transfers, activity intensity, and mobility that will maximize maintenance or return of function. Physical therapists in the ICU not only specialize in the mobilization of patients but, must also have expertise in the care of complex medical conditions, interventions, medications, and life-saving equipment.

The physical therapist in the ICU plays an important role in prevention and reduction of secondary conditions due to extended bed rest. High-level clinical decision-making is required to accurately dose physical therapy interventions and safely and effectively adjust treatment according to moment-to-moment changes in the patients’ hemodynamic stability and responses to movement and position changes. When mobilizing critically ill patients, physical therapist’s in the ICU must be able to synthesize information rapidly and accurately, integrating complex hemodynamic monitoring equipment in the context of the patient’s health, and medical management of the condition. Patients in the ICU have medical conditions that can change quickly. Physical therapists in the ICU are experts in identifying abnormal physiological responses to mobilization allowing professional adjustment of interventions to physiologic changes, while also managing life-supporting equipment. Mobilization becomes a very individualized and response-driven technique to allow for moment-to-moment evaluation and modification as needed.

A patient’s clinical status can change quickly in the ICU. Intensive-care unit acquired weakness can occur within 48-72 hours of ICU admission. Fluid shifts can be seen in as few as 3 days resulting in orthostasis, decreased stroke volume, and impaired cardiac output.
Given these rapid changes, specific outcome measures are warranted to identify deficits along the continuum of the patient’s ICU stay. In addition to the ICU physical therapist’s clinical reasoning, a robust outcome measure can help to guide decision-making, intervention prescription, and intensity.

**Physical Function Outcome Measures in the ICU**

Outcome measures provide information on whether goals are being met and whether interventions are effective. Standardized outcome measures provide a common language for practitioners and are the foundation for determining best clinical practices. Currently, there are no commonly accepted definitions for physical outcome measures. Researchers and current outcome tools available use terms such as “outcome measure,” “physical function,” “physical assessment,” or even “functional assessment,” making identification of all available measures difficult during literature searches. The lack of consistent terminology to define physical function makes it difficult to delineate between instruments that directly measure physical function versus other health concepts such as quality of life or cognition. Additionally, current outcome measures generally focus on impairment limitations rather than physical performance in relation to quality of life.

It is well-established that medical diagnosis alone cannot predict services needed nor provide enough information for health planning and management purposes. Patient survival is also not an adequate measure for patient centered outcomes. Chronic diseases have become overly prominent with more than 80% of the health care resources being dedicated to research and management. However, optimal clinical practice isn’t always translating to optimal clinical outcomes. Quality of life after discharge becomes a focus of the patient as well as the practitioner. Improvement, however, can only be demonstrated through direct measurement.
“The concept of quality of life is complex, and it embraces many characteristics of the social and physical environments as well as the health and internal states of individuals.”41 Two approaches towards quality of life include the subjective approach, patient perception; and the objective approach, external judgements on function relating to quality of life.41

Two resources that assist the physical therapist in the identification and definition of functional limitation are the World Health Organization (WHO) and the Patient Reported Outcomes Measurement Information System (PROMIS). The WHO has developed the International Classification of Functioning, Disability, and Health (ICF) model,23,35,39 which lists three major areas for inclusion in determining a person’s overall status: body functions and structure, activity, and participation restrictions.35,38,39 Within PROMIS, physical function is defined as the “ability to carry out various activities that require physical capability, ranging from self-care (basic activities of daily living (ADL) to more rigorous activities that require increasing degrees of mobility, strength, or endurance.”23,42 For the purpose of solidifying terminology that supports the definitions for both the WHO and PROMIS for physical function, this dissertation will utilize the term physical function assessment measure.

Given the focus of PT is on preventing or reversing the negative sequelae associated with bed rest while looking towards discharge recommendation from the facility and improvements in quality of life, physical therapists need to assess and monitor basic functional mobility skills. These skills often include bed mobility (including rolling, bridging, supine to/from sit), functional balance (sitting and standing), transfers, gait, and stairs.36,43 While body function/structural impairment measures can generate objective values, they may not provide the clinician full understanding of the patient’s functional status.37 A comprehensive physical function assessment measure can identify body function/structural impairments, activity
limitations, and participation restrictions that impact quality of life.\textsuperscript{44} Functional measures need to contain constructs that can assist with prediction of recovery and resumption of family, societal, and community roles.\textsuperscript{23}

For example, common ICU measures for body function/structural impairments are the medical research council sum score (MRC-SS) that measures strength deficits and incentive spirometry (IS) that evaluates lung function. While both reported outcomes give discrete values that can demonstrate deficits in strength and lung function, they cannot provide an accurate picture of whether a patient can walk to the kitchen or stand up from a chair. Values for the impairment outcomes may be low or abnormal, but a second functional outcome measure would be needed for complete evaluation of the patient with synthesis of the two optimal. Outcome measures that evaluate functional status, for example, ability to stand, perform ADLs, or to ambulate; would relate better to overall quality of life but may not offer the exact picture as to which specific impairment is causing the limitation.\textsuperscript{44} The clinician may still need to look at specific strength or respiratory impairment measures to identify the limiting factor for poor ambulation progression, thus requiring multiple outcome measures to be used. An outcome measure that combines both impairment and functional status assessments, providing more information regarding carrying out specific activities, abilities needed to function at home, and participate in life experiences; appears to be warranted.\textsuperscript{18,25,35,43}

The rehabilitation field has struggled as a whole for a comprehensive and yet clinically sensitive outcome measure for the ICU setting.\textsuperscript{23,35,37,43} An added challenge has been in designing a tool that has strong psychometric properties and is easy to use in an ICU setting.\textsuperscript{18,35,37,44} Since the aims of a physical function assessment measure are to provide rapid and accurate identification of impairments and functional performance that could delay recovery
during the ICU stay, this physical functional assessment measure should have at least 3 key psychometric properties: (1) validity, (2) reliability, and (3) responsiveness.\textsuperscript{18,44} Validity includes face, content, construct, and predictive.\textsuperscript{18,35} Reliability includes inter-and intra-rater reliability as well as test-retest reliability.\textsuperscript{18,44} Responsiveness of the measure is equally important given the moment to moment changes associated with ICU patients.\textsuperscript{25,30,31}

In the last three decades, more than 400 studies have been published tracking outcomes after critical illness.\textsuperscript{45,46} The limitation of all these studies is the lack of ability to synthesize the results.\textsuperscript{18,43,46} More than 250 outcome measures have been utilized throughout these published studies, making interpretation difficult and cohesive recommendations absent.\textsuperscript{46} Despite understanding the severity of illness and the high risk of mortality associated with patients who have an ICU stay, there is little consensus on utilization of ICU care or quality of care.\textsuperscript{47} A cornerstone concept for quality improvement, and changing outcomes post critical illness, is in preventability.\textsuperscript{47} Currently there is not a single outcome measure that covers the full spectrum of function, which limits a physical therapists ability to identify change and adjust their interventions accordingly. The need for multiple outcome measures or even changing outcome measures during a single stay can limit interpretability and introduce confusion and error. A single comprehensive physical function measure is needed not only to assist the physical therapist with clinical decision-making, but also to guide clinical research and recommendations for care.

**Statement of the Problem**

The Rehabilitation Measures Database currently reports more than 100 functional outcome measures.\textsuperscript{18,35} Utilization of an outcome measure in the ICU setting, however, is limited due to lack of specificity to the critical care population.\textsuperscript{18} Since critical care patients are a
heterogenous group with regards to age, acuity, and impairments, a generic physical function measurement tool is needed that covers a wide spectrum of patients in the ICU from medical to surgical, or cardiac to neurological.\textsuperscript{44} Measures developed for other settings and/or patient populations may not be valid for use in the ICU setting or critical care patient populations.\textsuperscript{37}

A second area of concern for outcome measures that are currently used with patients in the ICU setting is whether psychometric properties have been established for these measures.\textsuperscript{37} Parry et al.\textsuperscript{23,37} identified 26 measures used in ICU clinical research. The most common measures utilized were the Katz ADL and the 6-minute walk test. Neither of these measures has had psychometric testing conducted for use in the ICU setting. Six measures out of the 26 identified by Parry et al. were specifically developed for the ICU, but have very limited psychometric testing.\textsuperscript{23,37} Measures that are not valid or reliable can make results of a study not only meaningless but also potentially dangerous.\textsuperscript{18} Inaccurate clinical pictures can be painted by measurement systems that have not been fit to their purpose.\textsuperscript{18} The quality of studies that lack valid measurement properties can be interpreted as sub-standard.\textsuperscript{18}

A recent systematic review\textsuperscript{36} identified 14 physical function assessment measures with psychometric properties evaluated in the ICU setting. However, only 9 of these measures have been validated, and only 10 of them demonstrated predictive validity or responsiveness.\textsuperscript{36} The Physical Function in Intensive Care Unit test (PFIT), ICU Mobility Scale (IMS), Functional Status Score in the ICU (FSS-ICU), and the Acute Care Index of Function (ACIF) are the only 4 measures that have been found to be reliable, valid and responsive. None of these measures cover a full spectrum of impairment as well as functional assessment.
Rapid identification of impaired physical function is needed to prevent potentially negative sequelae associated with an ICU stay. To achieve optimal outcomes, the physical therapist in the ICU must be able to rely on a responsive, reliable, valid, and relevant measure that can capture a full spectrum of disability from very dependent to independent levels of function. That measure must also allow for correct interpretation of results and confidence in findings. Having one measurement tool that identifies all areas of impairment and function, can enhance the effectiveness and efficiency of PT care by reducing reliance on multiple tools as well as the need to extrapolate or coordinate interpretation of multiple measure results.

**Relevance and Significance of the Study**

Physical therapists in the ICU require a comprehensive and clinically sensitive instrument that is appropriate to use with hemodynamically unstable patients as well as efficient to administer. Early physical therapy initiation not only reduces the negative sequelae of bedrest but can reduce length of hospital stay and overall costs. Current study results for efficacy of treatment in the ICU, however, are varied due to inconsistent reporting of outcomes and lack of standardized care.

This study will address this deficiency by exploring and assessing current physical function assessment measures, and ultimately aggregate them into one comprehensive measurement tool that can be used specifically to assess patients in an ICU setting. Considering the limited energy expenditure possible for a patient in the ICU, and the need to identify the patient’s functional status efficiently and effectively, it is important to define the physical tasks needed to avoid redundancy through such a measurement tool.
Functional skills, including bed mobility, functional balance, transfers, gait, and stairs, allow for identification of a patient’s functional status. However, within the ICU there are complicating factors that may limit progression of function and prevent a safe discharge home. Outcome measures used within the ICU need to be feasible for the critically ill population. The measures used may also need to assess impairments not generally considered when assessing a patient in a rehabilitation or outpatient setting. Cognition, respiratory status, pain, hemodynamic stability and the presence of lines or tubes can dictate patient care and may hinder physical progression of the patient. Inclusion of these factors in an outcome measure could provide a holistic view of the patient’s function from a wide perspective.

Physical tasks performed on patients within the ICU must limit unnecessary testing and movement. For example, the de Morton Mobility Index (DEMMI) assesses bridging, rolling, and supine-to-sit for bed mobility. The standing balance category of the DEMMI has six components, which in includes jumping. However, requesting that a critically ill patient jump may not be appropriate. The argument can also be made that clinicians do not have to assess balance 6 ways to obtain an accurate functional picture, in turn, conserving the limited energy level of these patients.

When looking at impairment testing, specifically strength testing, there are 4 measures that currently assess strength. First, there is the MRC-SS, which assesses 6 muscle groups bilaterally for a composite score and has been highly correlated with identification of ICUAW. The remaining three measures include different components of a strength assessment. For example, the CPAx assesses grip strength, the PFIT is limited to quad strength and shoulder flexion strength, and the Perme relies on straight leg raise and shoulder flexion strength. It is
unclear if the full MRC-SS is needed within a physical function measure, as there is no consensus or study that supports which muscle groups can be substituted for the whole.

This study has identified the impairments and functional tasks that are needed to provide a comprehensive physical function assessment measure for ICU patients that captures a full spectrum of physical assessment from dependent to independent. This study utilized a group with broad acuity levels and multiple diagnoses to support generalizability across different health care ICU settings. Such a measure can provide efficient and effective identification of functional decline and better assist the physical therapist in clinical decision-making, geared towards treatment and discharge planning.

This process was best served with the use of IRT Rasch analysis. Researchers have used IRT Rasch analysis to improve measurement accuracy and reliability, while also reducing administrative time and effort during clinical use of the tool.\textsuperscript{56} IRT Rasch analysis allows individual tasks to be evaluated in depth, facilitating ranking of tasks for difficulty and removal of duplicates. For example, the several balance tests noted in the DEMMI were evaluated for redundancy and removal. IRT Rasch analysis also helped determine which parts of strength testing could be completed and substituted for the whole MRC-SS.

The complexity of the ICU setting makes creation of a comprehensive tool challenging although IRT Rasch analysis made it possible to devise a comprehensive tool that fosters effective assessment of patients in an ICU. Combining individual tasks from the 14 measures listed previously, assessing individual patients with individual tasks, analyzing rank order, and reducing duplicates has created a comprehensive physical function measure for use with the patient in an ICU setting. Thus, the use of IRT Rasch analysis identified the most appropriate
balance activity or strength assessment required in a comprehensive ICU physical function measure.

The overarching objective of this study was to have one physical function assessment measure that can be utilized for a variety of patient populations in an ICU setting. The use of a robust physical function assessment measure for use in future clinical studies allows for determination of efficacy of treatment, comparison of studies across settings, and broader interpretation of results.

**Specific Aims**

**Aim 1:** Identify physical-function measures currently utilized in the ICU that have been psychometrically tested.

*Objective 1:* Conduct an intense literature search to identify current physical function measures utilized in the ICU.

*Objective 2:* Extrapolate psychometric properties from all studies identified as well as define demographics including age, setting, and population size.

*Objective 3:* Extrapolate all individual tasks listed in the physical function measures identified to create a pool of test items.

**Aim 2:** Analyze all measure constructs to determine redundancies and appropriateness for use in the ICU setting according to Rasch analysis and IRT.

*Objective 1:* Identify Infit/Outfit of items according to results

*Objective 2:* Create Wright Map and ROC curve to determine characteristics of the testing items.
Aim 3: Create a comprehensive, robust functional measurement tool for use with patients in the intensive care unit.

Objective 1: Administer the pooled list of items to patients in the ICU as an initial evaluation and then upon discharge from the ICU.

Objective 2: Use IRT Rasch analysis for data analysis to determine which tasks should be included in the final physical function outcome measure.

Objective 3: Determine the predictive validity of the tool.

Summary

There are currently 14 physical function assessment measures utilized in the ICU setting that have been tested psychometrically. Given the limited level of psychometric testing that has been conducted for use of these measures in the ICU, results for clinical studies in the ICU need to be interpreted with caution.\textsuperscript{18,43} With multiple physical function assessment measures available, comparison and standardization of care is limited due to inability to synthesize current research. Currently, multiple measures are needed to address the full scope of impairments and functional deficits typically associated with ICU patients. This increases time of administration and may cause repetition in tasks administered.\textsuperscript{35,43} A standardized physical function assessment measure has been shown to be warranted to solidify veracity in research and to promote efficacy in clinical practice.\textsuperscript{18,23,43}

Development of a physical function measure required creation of the instrument, reduction or addition of items, assessment of the tool within the target population, and any further post testing revisions indicated.\textsuperscript{57} Measures developed must be appropriate, feasible, have interpretability across the target culture, reliability, validity, and responsiveness.\textsuperscript{57} With the
development of a comprehensive physical functional measure for use with patients in an ICU setting, the physical therapist is better equipped to direct care from the onset through rapid identification of impairments, development of an appropriate plan of care designed to reduce length of stay, and, finally, to advise on appropriate discharge recommendations. Within the ICU, discharge planning is an evolving process, and family planning is a daily occurrence. Accurate and efficient assessment of the progression or lack of progression over time in a quantifiable manner is necessary to maximize patient outcomes and reduce cost of care.

**Definitions**

**Delirium:** an acutely disturbed state of mind that occurs in fever, intoxication, and other disorders and is characterized by restlessness, illusions, and incoherence of thought and speech.

**Hand-grip strength:** force applied by the hand to pull on or suspend from objects

**Hemodynamic stability:** means that a person has a stable blood pressure and consistent flow of blood through his body. Hemodynamics is a term used to describe the intravascular pressure and flow that is produced by the heart's contractions.

**Intensive care unit (ICU):** the department of a hospital that is designed and equipped for the monitoring, care, and treatment of seriously ill or injured patients. Also known as “critical care unit.”

**Intensive care unit acquired weakness (ICUAW):** clinically weak ICU patients in whom there is no plausible etiology other than critical illness.
**International Classification of Functioning, Disability, and Health (ICF) Model:** is a classification of health and health-related domains looking at functioning and disability of an individual at both the individual and population levels while factoring in the environment.

**Life-saving equipment:** life support refers to the treatments and techniques performed in an emergency in order to support life after the failure of one or more vital organs. Not limited to ventilators.

**Mortality:** the state of being subject to death

**Physical therapist:** are highly educated, licensed health care professionals who can help patients reduce pain and improve or restore mobility - in many cases without expensive surgery and often reducing the need for long-term use of prescription medications and their side effects

**Post-traumatic stress disorder (PTSD):** a condition of persistent mental and emotional stress occurring as a result of injury or severe psychological shock, typically involving disturbance of sleep and constant vivid recall of the experience, with dulled responses to others and to the outside world.

**Post-intensive care syndrome (PICS):** a collection of health disorders that are common among patients who survive critical illness and intensive care. The range of symptoms that PICS describes falls under three broad categories: physical dysfunction, cognitive dysfunction, and mental health problems.

**Patient-Reported Outcomes Measurement Information System (PROMIS):** is a set of person-centered measures that evaluates and monitors physical, mental, and social health in adults and children utilized with the general population and with individuals living with chronic conditions.
**Quad extension strength**: ability to fully extend the knee against gravity

**Shoulder flexion strength**: ability to raise arm overhead, against gravity with arm fully extended.

**Straight-leg raise (SLR)**: ability to perform full range hip flexion with knee fully extended while lying supine in a bed.

**World Health Organization (WHO)**: is the directing and coordinating authority on international health within the United Nations’ system. WHO works together with policy-makers, global health partners, civil society, academia and the private sector to support countries to develop, implement and monitor solid national health plans, assure the availability of equitable integrated people-centered health services at an affordable price; facilitate access to affordable, safe and effective health technologies; and strengthen health information systems and evidence-based policy-making.
Chapter 2: Review of Literature

Introduction

Identification of limitations in physical function are important as limited physical function is related to decreased quality of life, increased risk of disability, falls, fractures and depression, as well as increase in healthcare costs. Physical function deficits and overall functional status can be determined through the use of an effective outcome measure. The use of outcome measures are a key component of an evidence-based approach to clinical decision-making that allows for quantifiable observations, communication across health-care settings, and a reduction to the potential variability of clinical recommendations, which may hold error and bias. Identification of an appropriate outcome measure to evaluate physical function requires that the instrument be “fit for purpose,” clinimetrically robust, and clinically applicable. This chapter identifies what is known and not known about physical function outcome measures currently utilized specifically in the ICU.

Historical Overview

Physical therapy is moving from the traditional fee-for-service or quantity-based care, to quality-based or value-based care. These changes have been initiated to support the three-part aim of better care for the individual, the population as a whole, and to lower healthcare costs. Treatment effectiveness must now be evident to patients, managers, employers, funders, physicians, and insurance companies. The Centers for Medicare and Medicaid Services (CMS), provide incentive payments for the quality of care for the medical services provided and have extended their value-based programs to settings in which PT is prominent. Current value-based programs include skilled nursing facilities and home health services. Acute care and outpatient physical therapy could quickly become part of this CMS reimbursement strategy, which should
drive physical therapists to utilize measures that can articulate the value they bring to the healthcare system.\textsuperscript{62,63} Transition from a collection of money for services rendered regardless of quality, to a system where payments are based on how well services benefit the patient, lower costs, and improve care, requires quality outcome measures.

Therapy practices have advanced significantly since the 1980’s when progression of care was subjectively described as “getting better,” “improving,” or “discharged.”\textsuperscript{64} The need for use of standardized outcome measures has been advocated by the World Confederation of Physical Therapy (WCPT), the Canadian Physiotherapy Association (CPA), as well as the American Physical Therapy Association (APTA).\textsuperscript{34,64} Standardized outcome measures are designed to measure various aspects of health status and fall under several terms such as “health status measures,” disability measures,” “outcomes measures,” or “quality of life measures.”\textsuperscript{65}

Outcome measures can assess the ability of the patient across the ICF spectrum: body structure/function, activity, and participation. Importantly, they also provide the practitioner a means to evaluate effectiveness of treatment.\textsuperscript{65} Outcome measures in physical therapy assist with professional accountability, transparency of the diagnostic process, clinical reasoning, and prognosis.\textsuperscript{66} The drive for use of standard outcome measures has been active for many years but studies revealed few physical therapists utilized them in their daily practice until CMS regulations changed in the United States (US) in 2017.\textsuperscript{67,68}

Early studies conducted in the 1990’s in Toronto, the United Kingdom, and Ireland reveal extreme variability in usage of outcome measures in inpatient-rehabilitation settings. In Toronto, only 20\% of therapists reported using an outcome measure.\textsuperscript{65} Of those 20\%, 88-90\% of physical therapists used manual muscle testing and goniometry as their outcome measure. Measures, such as the Functional Independence Measure (FIM), were at a low 18\% utilization, with therapists
reporting that their highest reason for non-use of an outcome measure was their discomfort with choices and interpretation. In the United Kingdom, 77% of rehabilitation centers surveyed reported use of at least one tool, only 28% of them, however, measured general motor function whereas 88% of them measured disability. In Ireland, only 30-50% of physical therapists utilized outcome measures.

By the 2000’s initiatives were made by many physical therapy professional associations - WCPT, CPA, APTA- supporting increased use of outcome measures. From 1998-2003, Ireland expanded their earlier study to not only include inpatient rehab centers, but hospitals, and outpatient centers. Authors Stokes & O’Neill demonstrated an increased use of outcome measures overall but found that very few disease specific tools were being utilized. Seventy-five percent of therapists surveyed reported use of standardized outcome measures, many endorsed by their facility. A study in 2011, looking at outcome measure usage in the Netherlands across inpatient rehabilitation, hospitals, and outpatient centers, revealed 70% of therapists utilized outcome measures. However, most of the measures used were directed towards pain, range of motion, and manual muscle strength. Again, very few therapists were utilizing measures of activity or participation.

In 2008, a survey by Jette and colleagues was conducted across the US to determine perceptions of outcome measure usage by physical therapists. One thousand physical therapists were randomly selected from the APTA database, with a final response of 456 usable questionnaires. The respondent’s demographics revealed 68% female, 61% worked in the outpatient setting, 53.4% had postbaccalaureate professional degrees, and 32% were certified clinical specialists. The results revealed that more than 50% of therapists did not use a standardized outcome measure. Although 90% of the respondents stated they agreed that
outcome measures can enhance communication and direct patient care, 70% of the respondents reported the reasons for non-use were: 1) too confusing for patients, 2) too difficult for patients to complete, and 3) too time consuming for patients. There was a higher use of outcome measures in the outpatient and home care settings than in the acute care setting where utilization of outcome measures was at 15%.65

Multiple review articles9,28,36,37,43,44,47,52,59 have been noted in the last two decades that address the concern and urgent need for a valid, reliable, comprehensive clinical physical function measure to address patients in the ICU. Outcome measures must be able to identify impairments but in a broader sense of how those impairments impact disability and functional status of a patient.70 The value in healthcare is dependent on improving performance and accountability while also enabling access to services, profitability, high quality, cost containment, safety, convenience, patient-centeredness, and satisfaction.71 Value becomes defined as health outcomes achieved per dollar spent. Adjusting services provided, decreasing staff or supplies to provide reduction in costs without regard to the outcomes being achieved limits effective care.71 When looking specifically at physical therapy services, justifying full-time staffing in the ICU, proper equipment for safe mobilization, and proper training in a high acuity setting, the outcomes must be documented to support the proposed changes. Without psychometrically sound outcome measures, it is difficult to measure value.72

In 2013, CMS mandated the use of functional reporting on all PT evaluations for any acute care in-patient who was being billed under Part B of Medicare.61 The purpose behind G-Codes was better understanding of beneficiary conditions, outcomes, and expenditures.61 Functional reporting consisted of 42 “G-Codes” with 7 severity/complexity modifiers.61 In January 2017, CMS released 4 new tiered billing code values for PT evaluations. Evaluations
must now be reported as “high complexity,” “moderate complexity,” or “low complexity.” The fourth billing code change was for re-evaluation of a patient. Although the billing codes themselves are tiered and must meet certain requirements, reimbursement was not tiered and remains the same for all levels. With the new complexity billing, patients must meet certain requirements across their history, examination of body systems, and clinical presentation. What must also be included within the evaluation is the use of an outcome measure. With the new CMS guideline, reporting is no longer an issue, rather quality of reporting is the mandate.

**Physical Therapy in the ICU**

Despite the variability noted in clinical trials, early mobilization and rehabilitation has been deemed to be safe and effective. A recent systematic review and meta-analysis confirmed a significant positive effect of PT in the ICU on the improvement in quality of life, physical function, peripheral muscle strength, respiratory muscle strength, as well as the ability to reduce overall length of stay in the ICU and the hospital. Mobilization of patients in the ICU starts with the acknowledgement of patient safety and hemodynamic stability. Physical therapy in the ICU has been shown to improve overall patient function, reduce ICU as well as hospital length of stay, and has been shown to be feasible and safe.

A recent systematic review encompassing 43 publications and more than 7,500 patients revealed a cumulative incidence of potential safety events in the ICU of only 2.6%. The authors also noted that overall medical adverse events in the ICU are common. The report indicated an incidence of 37% adverse events associated with morning care alone. This finding supports the need for the physical therapist in the ICU to have a knowledge base that can properly monitor the cardiac and respiratory stability of patients to provide accurate exercise prescription.
Ten published clinical trials have been reviewed, including 5 randomized control trials (RCT’s), one quasi-randomized trial, and 4 observational study.\textsuperscript{49,72} The variability in the use of outcome measures, timing and type of interventions, intensity and duration of care left cohesive recommendations difficult.\textsuperscript{49,72} The review also revealed, however, that current clinical studies leave more questions than answers on standard of care or outcomes due the lack of uniformity in outcome measures being utilized. The following provides a summative review of these 10 clinical trials.

In 2008, Morris et al\textsuperscript{77} conducted a Quasi-RCT in a single center Medical ICU looking at early initiation of physical therapy. Inclusion-criteria included patients who were within 48 hours of intubation and within 72 hours of admission. Patient’s had to be greater than 18 years of age and endotracheally intubated. Outcomes measures utilized were patients surviving to hospital discharge, days until first out of bed, number of ventilator days, ICU length of stay (LOS), and hospital LOS. The total number of participants was 330, split evenly between the intervention group and the usual care group. The intervention group focused on earlier initiation of physical therapy with out-of-bed first being recorded within 5 days of admission and the usual group within 11 days. No physical therapy specific interventions were changed or noted as being different between groups. Initiation of therapy day was the only difference between groups. The intervention group received at least one PT 80% of the time whereas the usual group received at least one PT session 47.4% of the time. Both groups were found to be statistically significant for both ICU length of stay and overall hospital length of stay being reduced by 2 days. The intervention group received an average of 5.5 sessions of PT as compared with the usual care group, which received 4.1, which was not statistically significant. There was no statistical significance noted between groups for ventilator free days or discharge disposition. There were
no reports regarding impairment or functional changes. Interventions were not discussed beyond their statement of standard care of passive range of motion or mobility therapy. This study was further limited by looking only at intubated patients.

In 2009, Schweickert et al. conducted a RCT in 2 different medical ICU settings looking at early physical and occupational therapy while patients were on mechanical ventilation. Inclusion criteria were patients greater than or equal to 18 years of age who had been on mechanical ventilation for less than 72 hours and expected to remain intubated for at least 24 hours. Patients had to have a baseline functional independence as defined by a Barthel Index score greater than or equal to 70 by patient or family self-report. One hundred and four patients were randomly assigned to groups. The intervention group consisted of initiation of PT and occupational therapy (OT) while the patient remained on mechanically ventilated. The standard at these facilities was that physical therapy and occupational therapy were not initiated while a patient was mechanically ventilated less than 2 weeks. PT and OT was initiated with the intervention group once inclusion criteria were met and the patients were progressed according to their individual ability from active assisted range of motion to out of bed or ambulation.

Schweickert et al. primarily utilized the FIM to measure the patient’s ability to return to full independent level of function; defined as the ability to perform six ADLs (bathing, dressing, eating, grooming, transferring from bed to chair, using the toilet), and walking independently. Secondary outcomes addressed were ICU and hospital LOS, number of ventilator free days, incidence of ICU-acquired weakness utilizing the MRC-SS, and handgrip strength upon discharge from the ICU. The Barthel Index was determined at discharge from the hospital, rather than from the ICU. The results of the study indicated that 50% of patients who received early PT & OT intervention were discharged home at an independent functional level. Statistical
significance between groups was noted in time of initiation to therapy, independence of ADLs at
time of hospital discharge, and greatest walking distance at time of hospital discharge. A
limitation in this study was that specific PT and OT interventions and progression of patients
were at discretion of the therapist.

In 2009, Burtin et al\textsuperscript{75} conducted a RCT in a medical and surgical ICU evaluating
benefits of cycle ergometry. Patients expected to have a prolonged ICU stay of at least 7 or more
days were eligible for inclusion. Judgment for readiness to begin PT was made on the fifth day of
admission to the ICU by the attending intensivist. The total number of participants for the
medical group were 90, the surgical group 71. The intervention group was initiated on cycle
ergometry 20 minutes, 5 days per week, the control group did not receive cycle ergometry. For
sedated patients, the ergometer was set to run passively. The intensity of the ergometer was set to
patient comfort for patients able to participate. The control and intervention group both received
the same standard of PT care, which varied from respiratory hygiene to passive range of motion,
out of bed, or ambulation for those who tolerated walking.

The primary outcomes assessed were the 6-minute walk test (6MWT) and the Short-Form
(SF)-36 physical function component, with both measured only upon discharge from the
hospital. Secondary outcomes were isometric quadriceps measured via hand-held dynamometry,
hand grip strength, and functional status measured via “sit-to-stand” ability. The results
indicated statistically significant improvement for the 6MWT, improved quadriceps force, and
improved SF-36 (PF) scores between groups but failed to demonstrate statistically significant
improvement in hand grip strength or “sit-to-stand” scores. The authors did not explain why the
intervention group was given an “extra” 20 minutes of treatment with cycle ergometry, which
might have influenced the outcomes. Another limitation of this study was the overall delay in
treatment or care, with the average onset of PT treatment at day 14 of admission for both the control and intervention groups.

In 2010, Routsi et al\textsuperscript{78} conducted a RCT at a multi-disciplinary ICU to evaluate the use of electrical muscle stimulation (EMS). Inclusion criteria was an Acute Physiology and Chronic Health Evaluation (APACHE) II admission score of 13 or greater within 24-48 hours of admission. The intervention group received EMS to the vastus lateralis, vastus medialis, and peroneus longus muscles of both lower extremities. The control group received usual care and no sham EMS. The total number of included participants was 142 with 70 being assigned to the intervention group and 72 to the control group. Due to many varied issues with completing the study, only 24 patients made it to the final for the intervention group and 28 patients for the control group. Outcome measures used were the MRC-SS and number of ventilator days.

Routsi et al\textsuperscript{78} found that critical illness polyneuropathy (CIPN) occurred in three of the intervention group patients and 11 patients of the control group as assessed by the MRC-SS. Days on the ventilator were calculated by median values, with reports of 2 days for the intervention group and 3 for the control group. However, both groups were noted to range from 0-99 days on the ventilator. The study failed to clearly define at what day the diagnosis of CIPN was made, leaving doubt as to whether this was bedrest acquired weakness or true critical illness myopathy. There was also no reporting of specific physical therapy interventions.

In 2013, Denehy et al\textsuperscript{79} conducted a RCT in a non-specified ICU setting. This was a longitudinal study looking at the intensity of PT services. A total of 150 participants were selected for the study. Inclusion criteria were patients within a 50-km radius of the hospital, no neurological, spinal, or musculoskeletal dysfunction that would prevent participation in physical rehabilitation, and an ICU LOS expected to be at least 5 days. Usual care was indicated as
mobility that may have included active bed exercises, sitting out of bed, and/or marching or walking. The intervention group received a specified prescription of intensity in all settings including ICU, hospital floor, and post discharge from the hospital. Within the ICU, the intensity of PT treatment was 15 minutes per day while on mechanical ventilation and 15 minutes twice per day once weaned from the ventilator. Intensity in the ICU and progression of care were tracked using the PFIT, the modified-Borg rating scale, and the SF-36. The 6MWT and the 5-repetition max test were conducted on the general hospital floor and as an outpatient with a 3-month, 6-month, and 12-month follow up.

Results of the initial 6MWT (on ICU discharge and admission to the hospital floor) demonstrated statistically significant difference between the intervention group and the control group, with the intervention group performing at a lower level than the control group. There was no statistical difference between groups for the PFIT within the ICU group, the modified Borg scale, or the SF-36. The authors indicated limitations of this study in the use of the 6MWT for patients at a low functional level. They also indicated that their facility does not routinely allow mechanically ventilated patients to move away from the bedside. More than 50% of their patients were mechanically ventilated at day 5 limiting their PT interventions.

In 2015, Moss et al\textsuperscript{80} conducted a RCT from 5 different ICU’s, at multiple medical centers to look at increased days of PT intervention. A total of 120 patients were selected from the inclusion criteria, which included patients who were at least 18 years of age and required mechanical ventilation for at least 4 days. The patients were randomized to either the intense PT group that received PT 7 days per week for at least 30 minutes or the standard of care group that received treatment at the discretion of the treating therapist for no more than 3 days a week. PT sessions in the ICU were scheduled for 30 minutes while sessions on the general floor, with
home care, or in the outpatient setting, were scheduled for 60 minutes. After hospital discharge to a home environment, the intervention group received continued protocol in the home or on an outpatient basis 3 days per week until the subject completed 28 days of therapy, or was able to successfully complete all stages of the program. Upon discharge, the control group received information on the importance of daily exercise, a follow-up phone call 3 times per week, and were encouraged to initiate their own exercise program. All patients were tracked for 6 months with phone calls and follow up.

Initial outcome measures utilized in the ICU were the MRC-SS, dyspnea scale scores, hand-grip strength by dynamometry, and the bed mobility portion of the FIM. The median initiation of PT was at 8 days post admission. The results demonstrated that patients who received the intensive therapy performed more standing activities than the control group, however this was not noted to be statistically significant. None of the patients in the intensive PT group achieved full independence upon discharge from the hospital. The 6-month follow up revealed the same results.

In 2013, Berney et al\textsuperscript{81} conducted a 1-day point prevalence study to determine mobility practices across 38 ICU’s in Australia and New Zealand. Data were collected on site regarding all mobility and rehabilitation activities undertaken by patients in the prior 24 hours. Data were collected from the nursing notes, physiotherapy notes, or from the daily observation chart, using forms created by the primary researchers. The two forms included a 30-item general case report form and a 25-item physiotherapy specific form. The study was performed at each site on one of three designated days in 2009 and 2010.

A total of 498 complete patient data sets were evaluated. Inclusion criteria was all adult patients (aged 16 years or over) who were admitted to the ICU at a 10am census time on the day
of the prevalence study. The study revealed the following practices: 28% of patients completed an in-bed exercise regiment, 19% of patients were able to sit up on the side of the bed, 37% of patients were able to sit out of bed into a chair, 25% of patients stood, and 18% of patients walked (distance was not reported). Overall findings reported no serious adverse events such as death, cardiac or respiratory arrest, or a patient fall. An interesting notation was that no patients on mechanical ventilation (n=200) were sat edge of bed or out of bed.

In 2014, Nydahl et al\textsuperscript{82} conducted a 1-day point-prevalence study to look at mobility practices across 116 ICU’s in Germany. The primary researchers created the data collection form used in the study, but the website link printed in the article was no longer retrievable. Data collection included airway type, highest level of mobilization achieved, most important barrier to mobilizing patients to a higher level (as perceived by the participating clinician), and most important complication (if any) occurring during mobilization (as perceived by the participating clinician). Inclusion criteria was all mechanically ventilated patients 18 years old or older currently admitted to an ICU. Practitioners who agreed to participate received e-mail notification by 7:00 am on the day prior to the official data collection day to gather the information from medical records. Participants had 3 days to complete data collection and enter it via the web-based form provided by the researchers.

A total of 783 complete patient data sets were evaluated. Mobilization out of bed occurred in 24% of all patients, with 55% having no mobilization greater than turning in bed and only 4% standing, marching, or walking on the day of the survey. Only 8% of patients with an endotracheal tube were reported to be out of bed and only 1 of 401 patients intubated was reported to stand, march, or walk. Mobilization out of bed did not differ by ICU type (i.e. surgical, cardiac, neurological), however, a greater proportion of patients (33%) were mobilized
out of bed in community and other hospitals as compared to university and university-affiliated hospitals. The reported difference was a higher frequency of complications at the university and university-affiliated sites.

In 2015, Jolly et al\textsuperscript{83} conducted a 2-day point prevalence study looking at mobility, which encompassed 42 specific ICU’s in the United States that were part of the Acute Respiratory Distress Syndrome (ARDS) Network. A standardized form was created by the primary research team and included the ICU Mobility Scale (IMS). Data were extracted on site from the medical chart at 8am on both days.

A total of 744 complete data sets were evaluated. Inclusion criteria was adults aged 18 years or older diagnosed with acute respiratory failure, requiring > 48hr of mechanical ventilation at any point during their ICU stay. Mechanical ventilation was defined as any ventilation via an endotracheal tube, tracheostomy tube, or noninvasive positive pressure ventilation, such as BiPap. Ongoing mechanical ventilation use was not required for eligibility. Study patients received mechanical ventilation on 73% of the patient-days. The prevalence of mobility provided by PT/OT was 32%. A significantly higher proportion of non-mechanically ventilated patients received PT/OT (48% vs 26%). Patients on mechanical ventilation achieved out-of-bed mobility 16% of the time. Mobilization without the involvement of PT/OT was 21%. PT/OT involvement in mobility was strongly associated with progression to out-of-bed mobility, whereas presence of an endotracheal tube and delirium were negatively associated with out of bed activities. Although it was indicated that the IMS was being utilized in the questionnaire, the specific scoring results of the IMS were not reported.

In 2015, Hodgson et al\textsuperscript{84} conducted a prospective observational study to look at early rehabilitation practices encompassing 12 ICU’s across Australia and New Zealand. A total of
192 patients were included in the study. Inclusion criteria was ability to independently mobilize prior to the current hospital admission, (includes gait aide, without assist of another person, or a wheelchair) in the ICU less than 72 hours, receiving invasive ventilation greater than 24 hours, and had an expected length of invasive ventilated for at least the next 48 hours. Patients received care according to standard hospital practices. The IMS and the MRC-SS were used as outcome measures. Findings noted that 63.5% patients did not receive early mobilization, which was defined as any active activity while the patient was on ventilation. Of the 36.5% of patients that did receive early mobilization; 45% were exercises in bed, 25% passively transferred to sitting, 11% sat at the edge of the bed, 5% stood at the bedside, 2% transferred from bed to chair through standing, and 12% walked. One-quarter of these patients were mobilized by day 3 and one-third by day 4 with the majority waiting until day 7 to ambulate if mechanically ventilated. A higher MRC-SS was associated with those that mobilized earlier.

A concerning limitation between the studies presented is the lack of one standard outcome measure that has been shown to be effective and efficient in the ICU setting. The studies varied usage of the 6MWT, Barthel index, MRC-SS, parts of FIM, PFIT, IMS, hand-grip strength, and/or hand-held dynamometry. When looking at these measures individually, one has to be cautious, for example the Barthel index and the 6MWT have not been validated with patients in the ICU. The RCT study by Denehy et al showed disappointing results with increased frequency of care, but perhaps use of the PFIT and 6MWT were not able to capture low-level patients in a high acuity setting. Generalized recommendations from these studies are difficult due to lack of consistency in outcome measures utilized that may not have clinimetric properties aligned with the specific measure and/or populations tested. The studies above do not appear to support the use of PT in the ICU given the limited positive results and lack of
statistically significant findings, however future studies could prove to be much different and improved with the correct usage of an appropriate outcome measure.

**Barriers to Physical Therapy in the ICU**

A recent survey was conducted in the US by Malone and colleagues\(^8\) to identify barriers to provision of PT in the ICU. The survey included questions and case scenarios that were developed by physical therapist clinicians and academicians, and critical care physicians. The final survey was a 65-item questionnaire set to a Likert scale and divided into two sections: Demographics of the hospital setting and the surveyed physical therapist, and the physical therapist’s perceptions on rehabilitation practices in the ICU. The survey was mailed to 2,320 physical therapists who were part of the Acute Care Section of the APTA. A response rate of 29% was reported encompassing 47 US States.\(^8\)

Similar to the point-prevalence findings by Nydahl et al.,\(^8\) this survey also found differences in patient care between community-based hospitals and academic or university type settings. Academic hospitals were noted to have lower ICU staffing than community hospitals with physical therapists working in the academic hospital having greater acute care experience. In contrast, physical therapists at the community-based hospitals reported a higher percentage of formal training (34.3%) but also a higher percentage of no training (14.3%) than those of academic institutions. Overall, 31.8% of physical therapists reported formal ICU training, whereas 55.9% reported informal hospital training, and 12.3% reported no training.\(^8\) Academic settings were more likely to have competency requirements than community-based hospitals. Sedation practices were highlighted as a barrier to mobilization more often in community settings than in academic settings. Forty-three percent of physical therapists reported termination of an ICU session according to specific department guidelines.\(^8\) Higher intensity interventions
such as out of bed or ambulation, were more likely to occur in academic hospitals than community hospitals.

Physical therapists in this survey, reported the medical complexity of the patient affected their decision-making and often their confidence toward progressive mobilization. Malone and colleagues encouraged initiatives directed towards increased awareness of the evidence supporting PT in the ICU, as well as to demonstrate the overall improvement in patient outcomes by having physical therapists present in the ICU.

**Outcome Measures for Physical Therapy**

Gaining access to psychometric information on outcome measures is easier since the creation of the Rehabilitation Measures Database (RMD). The RMD, created in 2010 by the Rehabilitation Institute of Chicago, is a free, online-access webpage for clinicians to obtain descriptions of available functional measures; their psychometric properties, clinical utility, and administration instructions. The RMD was developed through the use of focus groups and then followed by beta testing of the web page. A convenience sample of 75 rehabilitation professionals including nurses, physical therapists, and occupational therapists participated in 7 focus groups to provide recommendations for the database. Beta testing was conducted with 79 individuals from outpatient to inpatient settings including multiple disciplines.

The RMD currently provides information on more than 100 functional outcome measures being used in acute care and post-acute care. However, as noted above, the number of outcome measures that have been validated for use with patients in the ICU is limited. Measures developed and tested in one setting or patient population may pose threats to generalizability in usage. The role of an outcome measure is to assist with clinical decision-
making that enable observations to be quantified, comparisons made between sessions for
evaluation of progress, and communication and continuity of care between practitioners and
facilities to be enhanced. Outcome measures can show efficacy of practice, assist patients with
seeing their progress, and can facilitate insurance reimbursement. 59

Two recent studies have looked at perceptions, applications, and barriers to use of
standardized tests and outcome measures by physical therapists. 65,66 First, in 2008, Jette and
colleagues65 mailed 1,000 surveys to participants randomly selected from the APTA membership
list. The survey was designed by the investigators and the initial draft was sent to 14 colleagues
for additional input. The practice setting of these colleagues included acute care, outpatient, and
private practice. The purpose of the survey was to determine the extent of outcome measure
usage, perceptions regarding benefit and barriers to their use, and finally to examine factors
associated with their use. The response rate was 49.8%. 61% of the responses received were
from outpatient PT settings. 65

Survey respondents utilized standard outcome measures 47.8% of the time. More than
90% of respondents agreed that outcome measures enhance communication with patients and
direct their plan of care. The most frequent reasons for choosing a specific outcome measure was
ease of use, quick completion, and that the tool was valid and reliable. 65 Seventy-five percent of
respondents, however, reported that outcome measures were confusing for patients, difficult for
patients to complete, and too time consuming. Forty-nine percent of respondents indicated they
had no plan for use of an outcome measure due to time constraints imposed on both the patient
and the clinician; difficulty in analyzing, calculating and scoring; and finally, difficulty in
patients completing them independently. 65 Physical therapists working in outpatient settings or
home care settings were 7 and 12 times more likely to use an outcome measure than those in acute care.

In 2011, Swinkels and colleagues reported on current use and barriers to use of outcome measures in the Netherlands. In 2007, the Royal Dutch Society for Physical Therapy (KNGF) made active implementation of measurement tools and clinical practice guidelines a key aspect of its quality policy. While more than 18 clinical practice guidelines had been published and developed in the last decade, Swinkels and colleagues found the lack of standardized measures was limiting full guideline adherence.

The Swinkels and colleagues study was conducted in 3-phases. First, a general literature search on use of outcome measures in PT and perceived barriers was conducted. Second, a semi-structured interview of 10 physical therapists in the private sector and 10 physical therapists in nursing homes was used to identify facilitators and barriers to use of outcomes measures and determined actual usage of these measures. Finally, a survey was created by the researchers utilizing topics noted during their interviews and mailed to 2900 physical therapy members of the KNGF. The response rate was 16%.

A difference of the Swinkels et al study compared to the study conducted by Jette and colleagues was that respondents were asked to list what outcome measures they used. Respondents working in nursing homes listed 18 total measures used, with 5% of the measurement instruments mentioned once. In the private sector, more than 144 different measures were listed, however 58% of the measures were only listed once. The 6MWT appeared in both the nursing home (#2) and private sector (#5) as one of the top-5 most frequently used outcome measures. The private sector reported their first-place outcome measure was the visual analog scale for pain and the second-place outcome measure was goniometry. Both measures are
impairment based and quick/easy to perform. The nursing home therapists utilized the Berg Balance Scale as their number one and the 6MWT as their second. This article does not specify how many private practice therapists were from the hospital-based setting versus an outpatient clinic.

The two studies together\textsuperscript{65,66} indicated a list of barriers to the use of outcome measures. This list includes (in no specific order):

1) Administration, scoring, and interpreting results
2) Lack of administrative support and resources
3) Lack of financial compensation
4) Lack of familiarity or training with said measures
5) Lack of knowledge of psychometric properties
6) Support of colleagues in use of measurement instruments
7) Length of time needed to administer the measure
8) Inability for patients to complete them independently
9) Lack of time for identifying a suitable measure
10) Lack of agreement on which measure to use

The limitation with rehabilitation in the acute care setting and the ICU specifically, is the lack of clear consensus on the most important outcomes or measurement instruments.\textsuperscript{35,70} Despite the growing number of clinical studies within the rehabilitation field, few comparisons can be made and many studies cannot be synthesized due to the multiplicity of varied outcome measures being used.\textsuperscript{46} This would seem to make reasons number 5 – \textit{lack of knowledge of psychometric properties}, and 10 - \textit{Lack of agreement on which measure to use}, priorities as noted by both Jette et al\textsuperscript{65} and Swinkels et al.\textsuperscript{66} Identification of a suitable outcome measure from the
many that are available and agreement that this is the best measure for the current population appears to need further investigation.

Having a physical function outcome measure that is specific to a given patient population, efficient to use, easy to interpret and score, while also being valid, reliable, and responsive to track efficacy and quality of care are lacking for use in the ICU. Current mandates for use of an outcome measure, per CMS\(^6\) regulations, will alleviate the overall lack of utilization of an outcome measure in patient reporting. The new governing law, however, will not alleviate the barriers still present in many of the current measures being utilized.

**Use of Physical Function Outcome Measures in the ICU**

Measurement of health outcomes is essential in scientific research and in clinical practice for benchmarking performance and improving patient outcomes.\(^4\),\(^8\) Scores obtained with outcome measures enable decisions to be made about diagnostic tests and treatments.\(^8\) While there are several physical function measures readily available, many provide an incomplete picture and may not be adequate to drive quality of care or meaningful improvements.\(^4\) Health status measurement instruments need to be reliable, valid, and appropriate for the relevant patient population to avoid the risk of imprecise or biased results that might lead to wrong conclusions.\(^2\),\(^8\)

Physical function measures often used in the ICU by therapists and researchers include the 6-minute walk test (MWT), the Katz Activity of Daily Living (ADL), the Barthel Index, and the Functional Independence Measure (FIM). The 6-MWT does have clinimetric findings of good responsiveness, inter-rater reliability, and content validity with community-dwelling adults, outpatient cardiac and pulmonary rehabilitation centers, and patients with heart failure.\(^3\),\(^7\) The
6-MWT has also been used to evaluate patients post discharge from the ICU (once on the general hospital floor) and in follow up after discharge from the hospital. The Katz ADL has construct validity with the short-form 36 physical and mental domains but has no correlation with the FIM. The Katz ADL has been noted to have predictive ability of short-term mortality but not long-term mortality. The Katz has been used in the rehabilitation setting as well as in the post-ICU/acute-care setting.

The Barthel Index measures a patient’s ability to perform 10 basic activities of daily living including, feeding, grooming, bathing, dressing, bowel/bladder care, and toilet use, transfer to a chair, ambulation, and stairs. It has been assessed extensively in the stroke and geriatric populations and has been shown to be valid, reliable and responsive. The FIM has been extensively used in rehabilitation populations, evaluating 13 items of motor domains and 5 items of cognitive domains. The FIM looks at self-care, sphincter control, transfers, locomotion, communication, and social cognition. The FIM has been shown to have excellent reliability and validity, but has also demonstrated high ceiling effects upon discharge from the rehabilitation center or even one year post injury for patients with a neurological injury causing moderate to severe impairment.

Despite the use of these functional measures in recent RCT’s and observational studies conducted in the ICU setting, the psychometric properties of the 6MWT, Barthel, FIM, and Katz have not been demonstrated for use with patients in the ICU setting. For example, the 6MWT may be too advanced for many patients that present at a lower level of function and may hinder otherwise positive results from studies with poor choice in functional measure being utilized. These measures have not had clinimetric properties assessed showing evidence of reliability, validity, or responsiveness when used within the ICU setting or with a patient
population that includes those who are critically ill.\(^{28,37}\) While patients that present with higher levels of function may benefit from the 6MWT, patients at a lower level of function may require use of the Barthel Index. Use of multiple tests within the same patient admission, again, limits cohesive recommendations and guidance for practice and optimal care.

To date, there have been 31 studies establishing psychometric properties for 14 physical function measures designed specifically for the ICU setting.\(^36\) These measures included: (1) PFIT, (2) Chelsea Critical Care Physical Assessment Tool (CPAx), (3) Perme ICU mobility score, (4) Surgical intensive care unit optimal mobilization score (SOMS), (5) IMS, (6) FSS-ICU, (7) ACIF, (8) DEMMI, (9) Short Physical Performance Battery (SPPB), (10) Early Functional Abilities Scale (EFA), (11) Functional Capacity Scale (FCS), (12) MRC-SS, (13) Hand grip strength and hand-held dynamometry (HHD), and (14) Functional Assessment for Burns (FAB). Specific characteristics by relevant study article, can be seen in Appendix A.

Each of these published studies were recently assessed for quality\(^36\) using the COnsensus-based Standards for the selection of health Measurements INstruments (COSMIN) criteria appraisal tool using the COSMIN 4-point scale.\(^{88,89,91}\) The COSMIN aims to improve the selection of health measurement instruments. The checklist was developed via a Delphi study that included a multidisciplinary, international collaboration.\(^{88,89}\) Articles reporting psychometric properties are graded along ten domains: the presence of itemized response theory (IRT) testing, internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, and responsiveness.\(^{88,89}\) The original COSMIN grading system was a very broad “yes”/“no”/“can’t tell” answering system. A recent article by Terwee et al\(^90\) recommended and created an upgrade to the COSMIN scale by converting it to a 4-point grading system consisting of a 4-point ordinal scale of poor, fair, good,
and excellent. The lowest score assigned for an individual item along the checklist is the study's final quality score.

The results of the COSMIN scoring for the 31 articles that reviewed the psychometric properties of ICU specific physical function measures revealed that the PFIT, CPAx, and FSS-ICU are the most rigorously tested for psychometric properties, while the Perme and the FAB are the least. The EFA, FCS, Perme, and FAB were specific to certain patient populations limiting generalizability across multiple ICU settings and patient populations. The EFA and FCS are utilized primarily for patients within the neurological population, the Perme for cardiovascular, and the FAB for burns. Only 6 out of the 14 physical function measures identified have minimal clinical difference (MCD) or minimal clinically important difference (MCID) established.

In looking at the COSMIN quality scores across the 31 studies, only 2 outcome measures, the PFIT and CPAx had studies in which their domains scored excellent, though, overall, these outcomes measures still ranged from Poor to Excellent across all domains within all those studies. The PFIT-scored had one study that scored excellent in the area of itemized response theory (IRT) testing, internal consistency, and structural validity. The CPAx had one study that scored excellent with content validity. The FAB, MRC-SS, and FCS each had one domain within the studies scored as Good with the remaining being Fair or Poor. The FSS-ICU, IMS, ACIF, and DEMMI varied between Fair and Poor for all domains across all studies. The Perme, SOMS, SPPB, and EFA scored Poor across all domains for all studies. Findings across the 31 studies are concerning as only 6 studies score above a Poor along all domains, including: Corner et al, 2014, Ślusarz et al, 2012, Hermans et al 2012, Hough et al 2011, Lee et al, 2012, and Smailes et al, 2013. Full COSMIN scores can be seen in Appendix B.
ICU Physical Function Outcome Measures

Physical Function in ICU Test (PFIT)

The PFIT was originally developed in Australia, by ICU therapists using what they defined as the “pragmatic approach.” This study by Skinner et al utilized an initial sample of 10 participants in the ICU that had a tracheostomy in place. It was designed to measure endurance, strength, cardiovascular capacity, and functional level of patients in the ICU. The findings are used to guide prescription of exercise intensity and to evaluate progression of mobilization. The PFIT initially included 5-items: sit-to-stand, marching on spot, repetitive shoulder flexion, shoulder flexion strength, and knee extension strength. The PFIT was later updated to the PFITs (scored) after Rasch analysis was used to convert the scoring system to an interval level and the repetitive shoulder flexion component was removed. This is the version that is currently used in practice. The PFIT has been psychometrically tested in the ICU for inter-rater reliability and responsiveness to change. The PFITs has been psychometrically tested in the ICU for responsiveness to change, construct validity, convergent validity, discriminate validity, predictive validity, MCID, and floor and ceiling effects. The studies conducted for the PFIT and PFIT(s) have COSMIN scores ranging from “Poor” to “Excellent.”

Denehy and colleagues conducted the Rasch analysis on the PFIT and looked at its validity, responsiveness and predictive utility. This study included patients with a variety of diagnoses, although the authors do not clearly state whether only mechanically vented patients were initially recruited. The sample size was large at 116 final pairs of data. Though the scale was converted to interval level with the use of Rasch analysis, limitations were noted in a high floor effect on admission, and high ceiling effects on discharge. The authors indicated the
measure may not cover enough low or high order tasks to cover the breadth of functional range in the ICU.

A study completed by Nordon-Craft and colleagues evaluated responsiveness and predictive capabilities of the PFIT but, like Skinner and Colleagues, only looked at mechanically vented patients. The sample size for the Nordon-Craft study was limited to 34 participants. This study once again reports high floor effects. The authors also recommend combining the PFIT with other measures such as the MRC-SS and grip strength for a complete picture in evaluating for ICUAW. They also note the need for additional use of higher functional tests such as the 6MWT to evaluate the patient beyond the ability to march in place.

The first limiting factor with the PFITs, is in the strength assessment. The strength assessment is limited to shoulder flexion and knee extension, which may not capture full-strength impairments given limited muscular assessment. The instructions that indicate the use of “maximum strength of either the right or the left,” are vague and leave room for bias. There is no indication as to why shoulder flexion and knee extension were the only muscle groups chosen for strength testing. The second limiting factor is the starting level of performing a “sit-to-stand” from a chair. There is no assessment of bed mobility for lower level patients and no assessment of gait or stairs for higher level patients. The authors recommend combining multiple measures to provide an accurate functional picture of the patient.

While clinimetric testing of the PFITs has been thorough, sample size was only adequate for one study, missing data are an issue with all three studies, two of the studies limit the population to patients who are mechanically ventilated, and the limited spectrum of functional assessment make early identification of ICUAW or lower level impairments impossible. Psychometric property specifics of the PFIT and PFIT(s) can be found in Appendix C.
Chelsea Critical Care Physical Assessment Tool (CPAx)

The CPAx was originally developed in England in 2013. The CPAx uses an ordinal scale to assess 10 components of medical and functional assessment for ICU patients including: respiratory function, coughing, moving in the bed, supine to sitting edge of bed, dynamic sitting, standing balance, sit to stand, transfer from bed to chair, stepping, and grip strength. The CPAx was developed with a modified Delphi technique and later pilot tested across three different ICU settings. Initial testing included a very small sample size with 3 patients utilized to test for inter-rater reliability and 26 for construct validity. A follow up study conducted by the same authors in 2014, included 499 patients. The range of ICU patient diagnoses, however, was limited as they excluded cardiothoracic, burns, and neurological patients. A more recent study in 2015, examined patients with burns specifically, but the sample size was limited to 30 patients. The CPAx has been psychometrically tested in the ICU for inter-rater reliability, responsiveness, face validity, content validity, construct validity, floor and ceiling effects, and MCID.

COSMIN scores for the CPAx studies range from “Poor” to “Excellent.” The CPAx measures many of the lower level functional tasks and attempts to incorporate medical status with the inclusion of respiratory function and coughing ability. The CPAx, however, is limited in the “stepping” category because of vague wording. For example, the wording in Level 2-Stepping states Using mobility aids and assistance of at least one person (moderate),” Level 3-Stepping states “Using a mobility aide AND assistance of one person (minimal),” while Level 4-Stepping “Using mobility aide OR assistance of one person (minimal).” The wording of Level 5 Stepping simply states “fully independent.” There is little discrimination for progression of gait. There is no mention of distance walked, an assumption that maximum assistance will fall
under mechanical hoyer lift assistance, and there is no category for patients at a level requiring supervision. The second area of concern is in strength assessment, which is limited to grip strength.

The CPAx has a floor effect of approximately 67% on admission with the burn population, but reports a 3.2% floor effect for those with a general ICU admission. The benefits to the CPAx are the inclusion of medical status and respiratory status. The area most limited with use of the CPAx is in strength assessment. While psychometric testing has been thorough, reliability, internal consistency, and responsiveness have been found to be Poor. Psychometric property specifics of the CPAx can be found in Appendix C.

**Perme ICU Mobility Score (Perme)**

The Perme Score was developed by a physical therapist and first published in the United States in 2014. The Perme Score is an ordinal scale that considers mental status, potential mobility barriers, functional strength, bed mobility, transfers, gait, and endurance. The Perme Score was finalized by internal facility physical therapists, physicians, and nurses. No external review such as a Delphi study was completed. The Perme Score has received limited psychometric testing and was developed specifically for patients in a cardiovascular ICU.

There have been two studies of the Perme Score conducted to date and both were to determine inter-rater reliability. The two studies conducted by Perme et al & Nawa et al took place in the same facility, same patient population, and reported low sample sizes of 35 and 20 patients. Though the premise of the measure is promising, full psychometric testing is lacking and the measure would need to be generalized across different ICU settings as current testing was limited to use in a cardiovascular ICU. COSMIN scoring for the Perme ICU
Mobility Score studies were “Poor.” The strength of the Perme is its inclusion of mobility barriers and mental status that are often overlooked in other measures. Psychometric property specifics for the Perme Score can be found in Appendix C.

**Surgical Intensive Care Unit Optimal Mobility Score (SOMS)**

The SOMS was developed in the United States and first published in 2011. It uses an ordinal classification scale to assess daily mobility level achieved by patients in the ICU setting. The SOMS was initially designed as part of a mobility-start initiative to identify if mobilization of patients differed when being performed by nursing versus physical therapy. Garzon-Serrano and colleagues postulated there would be a difference between professions and sought a scale to capture those differences. The scale was originally validated, via face validity, with nurses reporting it was easy to understand, use and score.

There are 5 items to the SOMS: 1) SOMS 0 no activity, 2) SOMS 1 passive range of motion and sitting upright in the bed, 3) SOMS 2 sitting up in a chair, 4) SOMS 3 able to stand twice with Min A and march in place, 5) SOMS 4 able to ambulate. The SOMS has received psychometric testing in the ICU for inter-rater reliability, face validity, predictive validity, and cross-cultural validity.

COSMIN scoring for the SOMS studies are “Poor.” Despite overall sample sizes being very adequate at 63, 98, 113, and 128 participants, missing data left inconsistencies as to how many patients were tested for inter-rater reliability in the Kasotakis et al. and Schaller et al. studies. The nurse to nurse agreement in scoring was noted to be 65.6%, and nurse to physical therapist was 51.4-57.1%.
The SOMS is limited in its simple categorical design. Responsiveness and a suspected high ceiling effect come into question as these have not been evaluated in the ICU setting. SOMS 4 assesses the ability to ambulate. There is no differentiation between quality of ambulation, distance, or assistance needed. For example, a patient walking 5 feet at maximum assistance would score the same as a patient ambulating 300 feet independently. It would be difficult to demonstrate functional improvement and progression with these generic categories. Psychometric property specifics for the SOMS can be found in Appendix C.

**ICU Mobility Scale (IMS)**

The IMS was developed in Australia and first published in 2014. It was created by an ICU clinical trials group, which included physicians, physical therapists, and nurses. Expert opinion was used to create the initial tool, and feasibility testing included 15 nurses and 15 physical therapists as a convenience sample to determine ease of use. Further testing for face validity included a questionnaire emailed to more than 100 critical care professionals, which was used to finalize the IMS. This tool uses a 10-point ordinal classification scale to measure activity across 10 items: (0) Nothing, lying in bed, (1) Sitting in bed and exercises in bed, (2) Passively moving from bed to chair (i.e., Hoyer lift), (3) Sitting edge of bed, (4) Ability to stand (can use a tilt table), (5) transfer bed to chair (must take steps), (6) marching on spot (much take 4 marches), (7) walk with assistance of 2 or more people at least 5 yards, (8) walk with assist of 1 person at least 5 yards, (9) walk independently at least 5 yards with a gait aide but no assist, (10) walk independently no gait aid, no assist at least 5 yards. The psychometric properties of the IMS has been examined in the ICU for inter-rater reliability, face validity, content validity, convergent validity, predictive validity, and floor and ceiling effects. COSMIN scoring for the IMS studies are “Poor” to “Fair.”
While the IMS attempts to add more functional categories than the SOMS, it is similar to the SOMS in having been tested with both nursing and physical therapists. Despite the attempts to discern greater functional progression than the SOMS by broadening the ambulation related items, very high floor effects were noted in one study. The low numbers associated with the ceiling effect may lie in the use of the word “independent,” determined by the distance of 5 yards likely being achieved by subjects. Further studies need to be conducted with larger and more diverse populations. Psychometric property specifics for the IMS can be found in Appendix C.

**Functional Status Score for the ICU (FSS-ICU)**

The FSS-ICU first appeared in a pilot project in the United States in 2010. Zanni and colleagues report creation of the tool mirroring the FIM, which includes 18 functional activities. Three of those functional activities were deemed appropriate for use in the ICU setting by Zanni et al: bed mobility, transfers, and ambulation. The 5-item ordinal scale FSS-ICU includes: rolling, supine to sit, sitting edge of bed, sit-to-stand, and ambulation. All tasks are scored similar to the FIM with the Likert scale of 1 (complete dependence) to 7 (complete independence). The FSS-ICU was explored within the Long-Term Acute Care (LTAC) setting by Thrush et al for discharge predictability. Ragavan and colleagues looked at the measure for inter-rater reliability and discharge predictability from the ICU. The ICU study was limited by a small sample size of 26 patients. Parry et al most recently looked at the FSS-ICU in comparison with other ICU functional measures using a larger sample size of 66 participants, offering convergent validity, MCID, and floor and ceiling effects. COSMIN scoring for the FSS-ICU studies are “Poor” to “Fair.”

There was a recent international clinimetric study completed regarding the FSS-ICU in 2016. Data for this study were pooled from 5 international databases. Missing data made
extrapolation of information difficult. It is unclear how many patients were specifically from the ICU. There is mention of pre-hospital admission and post hospital discharge FSS-ICU scores, but it is unclear what data were taken during the ICU stay itself. Two studies mentioned, for which data were extrapolated, were not referenced by the author limiting ability to review the studies personally to obtain the specific information needed.

The second limitation in the FSS-ICU, specifically, is within the “FIM” scoring, where patients must achieve a certain level of assistance, or lack of assistance, as well as distance walked to advance in scoring. For example, the patient must achieve a distance of 150 feet (ft) walked with no more than minimal assistance to receive a score of “3.” There is concern as to responsiveness of the tool with use on patients who are at a low level of function. For example, a patient who ambulates 10 ft with maximum assistance of one person scores the same as a patient who ambulates 100 ft with supervision. There is no indication that the FSS-ICU can also use the “household exception” rule that is used with FIM scoring, which indicates instead of achieving 150 ft the patient can ambulate 50 ft. Responsiveness has not been clinimetrically tested in the ICU setting. Psychometric property specifics for the FSS-ICU can be found in Appendix C.

Acute Care Index of Function (ACIF)

The ACIF was developed in the United States and first appears in the literature in 1988. It was designed initially for patients with acute neurological impairment in the acute care setting. The ACIF consists of 20 items divided into 4 subset categories: Category 1: Mental Status - verbal commands, commands, learning, safety awareness; Category 2: bed mobility – rolling right, rolling left, supine-to-sit, sit-to-supine; Category 3: transfers – wheelchair to mat, mat to wheelchair, sit-to-stand, stand-to-sit, sitting balance, standing balance; Category 4:
mobility – gait with device, gait without device, ascend steps, descend steps, propel wheelchair, set up wheelchair. The ACIF has been psychometrically tested in the ICU in one study with a sample size of 42. Inter-rater reliability, convergent validity, and predictive validity were assessed. COSMIN scoring for the ACIF study is “Poor” to “Fair.”

The ACIF uses a weighted scoring system. More weight is given to ambulation and transfers than to the mental status or bed mobility categories. This emphasis may be quite different in ICU environments where patients may have more bed mobility and mental status deficits. Further assessment and testing of the scoring system are needed in the ICU with larger sample sizes with perhaps a change of the weighted system warranted for the ACIF.

The ACIF also includes the categories of wheelchair set up and ability to propel, which may not be appropriate with certain patient populations. The ACIF has been modified for orthopedic populations to exclude these items, but this version has not been assessed for use in the ICU. There is also concern with the grading system for each domain: “unable,” “dependent,” and “independent,” where a lack of delineation between assist levels or levels of supervision exist. A larger study needs to be conducted to look at overall responsiveness of this tool within the ICU setting, since many patients in the ICU might fall into the “dependent” category for a majority of their stay. A larger sample size might improve the accuracy in capturing small functional gains in a critically ill patient population. Psychometric property specifics for the ACIF can be found in Appendix C.

de Morton Mobility Index (DEMMI)

The DEMMI was initially created in 2008 in Australia to measure physical performance and mobility activities for older adults across diverse clinical settings. The DEMMI uses an
interval score to measure 15 items: bridging, rolling, supine-to-sit, sit unsupported in chair, sit-to-stand from chair, sit-to-stand without using arms, standing unsupported, standing feet together, standing on toes, tandem stand with eyes closed, walking distance with/without a gait aid, walking assistance, picking up a pen from the floor, walking 4 steps backwards, jumping.

The DEMMI has been tested for the ICU in one study for inter-rater reliability, intra-rater reliability, convergent validity, MCID and floor and ceiling effects. COSMIN scoring for the DEMMI studies are “Poor” to “Fair.”

The DEMMI stands out from other outcome measures in that its original creation was through Rasch analysis, which allowed for selective inclusion of test items. Second, it has psychometric testing for a diverse range of elderly patients with acute and chronic illnesses in the community, rehab, and acute care settings. However, questions can be raised regarding the feasibility of having critically ill patients jump or the number of balance tests needed for accurate assessment. Redundant assessment of tasks should generally be avoided in patients with limited energy reserve. Further testing needs to be completed with the critically ill population. Psychometric property specifics for the DEMMI can be found in Appendix C.

**Short Physical Performance Battery (SPPB)**

The SPPB originated through a United States study that took place from 1981-1984. The SPPB is a 3-task, timed based test designed to assess lower extremity physical performance. The three tasks are standing balance, walking speed, and chair stand tests. The balance tests include side-by-side standing, semi-tandem standing, and tandem standing. The walking speed test is calculated twice over 3 or 4 meters. The chair stand test consists of repeating 5-stands from a chair without using arms, twice. A recent systematic review and meta-analysis revealed that an SPPB score < 10 is predictive of all-cause mortality. Use of the SPPB in the ICU has
been studies once for clinimetric properties including convergent validity, predictive validity, MCID, and floor and ceiling effects.\textsuperscript{23} That study was limited by a small sample size of 23 participants. The COSMIN score for the SPPB study is rated as “Poor.”\textsuperscript{36}

The limitations for the SPPB is in the need for repetitive testing. This may be difficult for patient’s in the ICU setting given lines and tubes for distances walked but also due to limited energy available for this population to repeat tests multiple times.\textsuperscript{23} Several balance tests are required for the SPPB and standing from a chair without the use of arms can be considered difficult for patients in the ICU setting.\textsuperscript{23} The floor effect on both admission and discharge from the ICU were > 50\%, making consideration of this test low for this population. Psychometric property specifics for the SPPB can be found in Appendix C.

**Early Functional Abilities Scale (EFA)**

The EFA scale was first introduced in Germany in 2000.\textsuperscript{123} It is designed to address activities of daily living (ADLs) and cognitive function (including wakefulness) of neurosurgical patients. There are 4-categories that include 20 items to assess the patient’s neurologic function: (1) vegetative functions including autonomic stability, wakefulness, tolerance to postural changes, and continence; (2) oro-facial functions including oral hygiene, swallowing, tongue movements/chewing, and facial expression; (3) sensorimotor abilities including muscle tone, head posture control, trunk postural control, changing of position, standing, voluntary movements, and locomotion or wheelchair use; (4) cognitive abilities including tactile stimulation, visual stimulation, auditory stimulation, communication, and comprehension. Clinimetrics have been tested for the EFA that include inter-rater reliability and floor and ceiling effects in a small sample size of 24 patients.\textsuperscript{124} COSMIN scoring for the EFA studies have been rated as “Poor.”\textsuperscript{36}
The EFA limits itself for use to the traumatic brain injury (TBI) population. While the
cognitive assessment sections are more in-depth than any other current physical function
measure, the specific neurological assessment for vegetative state is much more intense. The
mobility section is lacking. For example, level 3 locomotion is ability to maintain postural
control while seated; level 4 locomotion is active use of wheelchair or “walking some steps;”
while level 5 locomotion is independent with wheelchair and walking. The EFA appears to focus
heavily on the cognitive domain, more so than the physical domain. The neurological focus of
the EFA limits its use across ICU settings. Psychometric property specifics for the EFA can be
found in Appendix C.

Functional Capacity Scale (FCS)

The FCS was published in 2003, however, initial studies and reports are in Polish and
outside the scope of this dissertation. Later studies were available in English translation. The
FCS was created as an easy functional assessment tool for the neurosurgical population. It is an
ordinal scale of 4 categories. Scoring is similar to the SOMS and the IMS. In Level 1, the patient
does not require any assistance/has independence; in Level 2, the patient requires assistance/has
moderate independence; in Level 3, the patient requires significant help/has moderate
dependence; and in Level 4, the patient requires intensive care/is dependent. While the FCS has
psychometric testing completed for the ICU including inter-rater reliability,\(^\text{93}\) convergent
validity,\(^\text{93,125}\) and predictive validity;\(^\text{126}\) all testing has been limited to specific neurosurgical
populations including TBI, brain aneurysms, and brain tumors. COSMIN scoring for these
studies was “Poor” to “Good.”\(^\text{36}\)

Given the categorical design, more testing is needed to generalize this across multiple
ICU settings, with the same arguments made as with the SOMS and IMS regarding simplified
categorical scales and limits to responsiveness. There is also subjectivity for Levels 2 and 3 in categorizing “needs assist” versus “significant help/moderate dependence.” This measure would not be recommended for use in the general ICU setting. Psychometric property specifics for the FCS can be found in Appendix C.

Medical Research Council Sum Score (MRC-SS), Grip Strength, and Hand-held dynamometry (HHD)

The MRC-SS evaluates the strength of three muscle groups on all 4 limbs using an ordinal scale that assigns a score of 0-5 to each muscle group to a perfect composite score of 60. The MRC-SS was first utilized in a 1988 pilot study of patients with Guillain-Barre. It has since been used to identify ICUAW. ICUAW is defined as having an MRC-SS less than 48 or “significant weakness.” There has been further delineation for values below 36 indicating “severe weakness.” Despite the apparent simplicity of testing, there are arguments between reliability of identification of ICUAW given variability in therapist testing, and with the feasibility of accurate testing in patients that may or may not be able to participate fully with manual muscle testing due to sedative effects in the ICU or ability to follow commands.

Within the ICU, the MRC-SS & hand grip strength along with use of hand-held dynamometry for muscle testing, has demonstrated inter-rater reliability, and predictive validity. MCID values have been found for the use of hand-held dynamometry for testing muscle strength. The COSMIN scores for these studies ranged from “Poor” to “Good.”

There are many arguments surrounding the use of manual muscle testing. Studies have looked into substituting hand grip strength for MRC-SS testing, with some arguing that manual muscle testing should be forgone in the ICU setting due to variability in testing.
inconsistencies, and patient participation.\textsuperscript{132} An additional argument accompanying the MRC-SS is the differentiation of grades 4 and 5, as previously noted. A third argument can be made that despite a lack of apparent weakness according to the MRC-SS, one might still suspect a weakness to be present relative to baseline functioning (i.e., ability to stand, walk, perform transfers). This would suggest caution be used with the MRC-SS as the only outcome measure. Given its popular use, it would be helpful to know if the MRC-SS could be utilized solely for specific movements such as hip flexion or shoulder flexion. More clinimetric testing and correlation is needed with specific items to determine the use of the MRC-CC in parts versus the whole. Psychometric property specifics for the MRC-SS and HHD can be found in Appendix C.

### Functional Assessment of Burns (FAB)

The FAB, published in 2013 in the United Kingdom, assesses burn patients in the ICU.\textsuperscript{96} The scale was created by physical therapy staff working in the hospital burn unit where the study was conducted. There are 7 tasks completed on the FAB on a 1-5 scale. The tasks are feeding, washing, toileting, transfers, dressing, walking, and stair climbing. The scoring is similar but not identical to the FIM. The scoring criteria includes: (1) if assessed to be fully dependent or the task is unable to be assessed at all; (2) patient completes the activity with physical assist; (3) patient completes the activity with supervision or verbal cueing; (4) patient completes the activity independently but with an aide; and (5) if the activity can be completed independently. Clinimetrically, the FAB has been evaluated for discharge predictability. The one ICU study for the FAB has a COSMIN score of “Good.”\textsuperscript{36}

As noted with other similar scoring measures, questions arise with regard to responsiveness and tracking progression. It is not clearly defined as to whether a Level (2) patient who ambulates 10’ with maximum assistance would score the same as a patient walking
100’ with minimal assistance. A patient may change his/her walking speed, distance, or assistance needed, and this would not be reflected in his scoring until he/she reaches a supervision or independent level. Limitations are present as this is a single study in which predictive validity is the only current clinimetric value. Additionally, the predictive validity has only been completed specifically with a burn population. Psychometric property specifics for the FAB can be found in Appendix C.

What is Known and Not Known

While 14 physical function measures with clinimetric properties have been identified for use in the ICU, it is not known what tasks are the most important to be included in an all-inclusive and feasible measure for patients in an ICU setting. Is standing marching better or equal to ambulation? Can a straight leg raise be substituted for the whole MRC-SS? Which balance tests are the best, or do we need to assess balance with multiple techniques? Each assessment measure has identified associated tasks to be included. Some are very heavy on balance assessment, while others do not assess strength. Bed mobility was a task assessed in only a few of the measures and stairs were often neglected. See Table 1 for physical function measure components.

Table 1: Physical Function Measure Components

<table>
<thead>
<tr>
<th></th>
<th>Strength</th>
<th>Bed mobility</th>
<th>Balance</th>
<th>Transfer</th>
<th>Gait</th>
<th>Stairs</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PFIT/PFITs</strong></td>
<td>-shoulder strength</td>
<td>-rolling to sit</td>
<td>-standing balance</td>
<td>-sit to stand</td>
<td>-sit to stand</td>
<td>-ability to take steps</td>
<td>step cadence</td>
</tr>
<tr>
<td></td>
<td>-quad strength</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CPAx</strong></td>
<td>grip strength</td>
<td>-rolling to sit</td>
<td>-standing balance</td>
<td>-sit to stand</td>
<td>-sit to stand</td>
<td>-ability to take steps</td>
<td>-cough</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>function</td>
</tr>
<tr>
<td>Perme Mobility Score</td>
<td>Strength</td>
<td>Bed mobility</td>
<td>Balance</td>
<td>Transfer</td>
<td>Gait</td>
<td>Stairs</td>
<td>Other</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>Perme Mobility Score</td>
<td>-straight leg raise</td>
<td>-supine to sit</td>
<td>-seated balance</td>
<td>-sit to stand</td>
<td>-assistance needed and distance walked in 2 minutes</td>
<td>-pain</td>
<td>-mental status</td>
</tr>
<tr>
<td>SOMS</td>
<td>-Sit edge of bed without support</td>
<td>-perform sit to stand twice with min A</td>
<td>-ability to ambulate</td>
<td>-steps in place</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS</td>
<td>any bed activity: rolling, bridging, therex</td>
<td>-sitting edge of bed</td>
<td>-transfer bed to chair</td>
<td>-varied gait qualifiers</td>
<td>-marching on spot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSS-ICU</td>
<td>-rolling -supine to sit</td>
<td>-sitting edge of bed</td>
<td>-sit to stand ambulation</td>
<td>-strongly based on FIM (can substitute WC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACIF</td>
<td>-rolling -supine to sit -Sit to supine</td>
<td>-sitting balance -standing balance</td>
<td>-chair to bed -bed to chair -sit to stand -stand to sit</td>
<td>-with device -without device (can substitute wheelchair)</td>
<td>-</td>
<td>-mental status</td>
<td></td>
</tr>
<tr>
<td>DEMMI</td>
<td>-bridge -rolling -supine to sit</td>
<td>-Sitting balance -standing balance -pick up pen from floor</td>
<td>-sit to stand -sit to stand without using arms -distance and assist backwards gait assessment</td>
<td>-ability to jump</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPPB</td>
<td>-standing balance</td>
<td>-sit to stand -5x sit to stand gait speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCS</td>
<td>-ability to walk</td>
<td>-personal hygiene -medical status -pain -mood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EFA</td>
<td>-sitting balance -standing balance</td>
<td>-Sit to stand -transfer to chair</td>
<td>-mental status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A small table showing different scales and their components:

<table>
<thead>
<tr>
<th>Strength</th>
<th>Bed mobility</th>
<th>Balance</th>
<th>Transfer</th>
<th>Gait</th>
<th>Stairs</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-swallowing &amp; chewing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-sensory</td>
</tr>
<tr>
<td><strong>FAB</strong></td>
<td></td>
<td></td>
<td>transfer</td>
<td>ability to walk</td>
<td>ability to climb</td>
<td>-feeding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bed to chair</td>
<td>10 meters</td>
<td>FF stairs</td>
<td>-washing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>transfer from toilet</td>
<td></td>
<td></td>
<td>-toileting</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-dressing</td>
</tr>
<tr>
<td><strong>MRC-SS/HHD</strong></td>
<td>strength assessment only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Physical Function in Intensive care unit Test (PFIT), Chelsea critical care Physical Assessment tool (CPAx), Perme ICU mobility score (Perme), Surgical intensive care unit Optimal Mobilization Score (SOMS), ICU Mobility Scale (IMS), Functional Status Score for the ICU (FSS-ICU), Acute Care Index of Function (ACIF), De-Morton Mobility Index (DEMMI), Short Physical Performance Battery (SPPB), Early Functional Abilities scale (EFA), Functional Capacity Scale (FCS), Medical Research Council Sum Score, (MRC-SS), Hand Held dynamometry (HHD), Functional Assessment for Burns (FAB), minimal (Min), assist (A), wheel chair (WC), full flight (FF)

*MRC-SS and HHD testing was combined as they are measuring the same construct.*

Given the complexity of the ICU, a patient’s functional status can vary greatly on a day to day basis. Having an outcome measure that can capture not only progression, but regression would be ideal. If an assessment measure starts with the ability to stand, capturing outcomes for patients functioning at a lower bed mobility level would be limited. Being able to track even small changes in function is essential for developing prevention strategies. There is also the argument for identification of ICUAW. Diagnosis of ICUAW is currently by clinical suspicion and the MRC-SS value. None of the assessment measures discussed in this chapter include the full MRC-SS unless it is completed as a secondary measure. There are some assessment measures that utilize specific areas of strength testing, such as, shoulder flexion and knee extension. Another uses grip strength, and another uses the straight leg raise. However, there is no evidence to support use of a few motions in place of all motions on the MRC-SS.
Assessment of Medical Acuity

Rehabilitation/Mobilization Criteria

A recent panel of 32 international experts, 4 methodologists, and 4 critical illness survivors collaborated over 6 years in person, via teleconferences, and via electronic communication to update and expand the 2013 Clinical Practice Guidelines for the Management of Pain, Agitation, and Delirium in Adult Patients in the ICU. Expansion of these guidelines included the section of mobilization/immobility, thus creating the 2018 Clinical Practice Guidelines for the Prevention and Management of Pain, Agitation/Sedation, Delirium, Immobility, and Sleep Disruption (PADIS) for Adult Patients in the ICU. The 2018 PADIS recommendations define rehabilitation as a set of interventions with mobilization at the core that optimizes function, reduces disability, and facilitates movement for patients with a health condition to improve patient outcomes.

The recommendations for initiation of interventions covered four specific areas including: efficacy and benefit, safety and risk, indicators for initiation, and indicators for stopping. With regard to efficacy and benefit, the recommendation was conditional for performing rehabilitation or mobilization in critically ill adults. This recommendation was based on the results of 16 RCTs, although evidence was reported to be low quality. For safety and risk, it was reported that serious safety events or harms do not occur commonly during physical rehabilitation or mobilization. Rationale for this recommendation came from 10 observational studies and 9 RCTs.

The indicators for initiation of intervention included cardiovascular, respiratory, and neurologic status. Vasopressor infusion and mechanical ventilation were not deemed barriers for initiation to mobilization. From a cardiovascular standpoint, the parameters to start mobilization...
included: heart rate 60-130 beats per minute (bpm), systolic blood pressure (SBP) between 90-180 millimeter of mercury (mmHg), and a mean arterial blood pressure (MAP) 60-100 mmHg. The respiratory parameters to start mobilization were respiratory rate 5-40 breaths per minute, pulse oximetry (Spo2) ≥ 88%, fraction of inspired oxygen (FiO2) < 0.6, and positive end expiratory pressure (PEEP) < 10. From a neurological status, the patient had to be able to open their eyes to voice to initiate intervention. The indications for stopping mobilization were parameters outside of these recommendations.133

The recommendations, however, also specifically stated that although these parameters were determined by clinical research and interpreted via expert opinion, they should not serve as a substitute for clinical judgement. “All thresholds should be interpreted and modified, as needed, in the context of individual patients’ clinical symptoms, expected values, recent trends, and any clinician-prescribed goals or targets.”133, pE850

The Richmond Agitation-Sedation Scale (RASS)

The RASS scale was developed by critical care physicians, nurses, and pharmacists to measure a patient’s level of consciousness or agitation.134 RASS is a 10-point scale with 4 levels denoting anxiety or agitation, one level indicating calm and alert, and 5 levels indicating sedation. See Table 2 for RASS scoring. A study by Sessler and colleagues, 2002135 evaluated reliability and validity of the RASS in adult intensive care units. The study utilized 2 physicians, 2 nurses, and 1 pharmacist and evaluated 172 patients. The ICU’s included medical respiratory, neuroscience, coronary, surgical trauma, and cardiac surgery. RASS scores were lower for patients who were mechanically ventilated (p < 0.0001), receiving continuous infusion sedative or analgesic medication (p < 0.001), or had higher Acute Physiologic and Chronic Health Evaluation (APACHE) II scores (p < 0.05). Inter-rater reliability was excellent across the entire
ICU population (ICC 0.956), ICU sub-groups (ICC 0.922-0.983), and pair-wise comparison between investigators (r = 0.944-0.973).  

A second study completed by Ely and colleagues in 2003 also looked at reliability and validity of the RASS scale with critically ill patients. This study was conducted by two nurses, one physician, and one neuropsychiatric expert. The study took place in the medical and coronary ICUs. The validity portion included 275 patients. The reliability portion included 38 patients. The RASS has criterion validity against the neuropsychiatric expert (p < 0.001), Content validity against the Glasgow Coma Score (r = 0.91, P < 0.001) and EEG recordings (r = 0.64, P < 0.001), and face validity via a survey of internal critical care nurses. Inter-rater reliability between all groups ranged from weighted k = 0.79-0.91.

Table 2: Richmond Agitation-Sedation Scale (RASS)

<table>
<thead>
<tr>
<th>Score</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4</td>
<td>Combative</td>
<td>Overtly combative or violent; immediate danger to staff</td>
</tr>
<tr>
<td>+3</td>
<td>Very Agitated</td>
<td>Pulls on or removes tube(s) or catheter(s) or has aggressive behavior toward staff</td>
</tr>
<tr>
<td>+2</td>
<td>Agitated</td>
<td>Frequent non-purposeful movement or patient–ventilator dyssynchrony</td>
</tr>
<tr>
<td>+1</td>
<td>Restless</td>
<td>Anxious or apprehensive but movements not aggressive or vigorous</td>
</tr>
<tr>
<td>0</td>
<td>Alert and Calm</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>Drowsy</td>
<td>Not fully alert, but has sustained (more than 10 seconds) awakening, with eye contact, to voice</td>
</tr>
<tr>
<td>-2</td>
<td>Light Sedation</td>
<td>Briefly (less than 10 seconds) awakens with eye contact to voice</td>
</tr>
<tr>
<td>-3</td>
<td>Moderate Sedation</td>
<td>Any movement (but no eye contact) to voice</td>
</tr>
<tr>
<td>Score</td>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>-4</td>
<td>Deep Sedation</td>
<td>No response to voice, but any movement to physical stimulation</td>
</tr>
<tr>
<td>-5</td>
<td>Unarousable</td>
<td>No response to voice or physical stimulation</td>
</tr>
</tbody>
</table>

**Acute Physiology and Chronic Health Evaluation (APACHE) II**

Diagnosis is not enough to quantify the severity of illness. The APACHE score was designed to calculate three patient factors that influence acute illness outcome. The APACHE looks at pre-existing disease, patient reserve, and severity of the acute illness. The original APACHE was released in 1981 and updated to the APACHE II in 1985 reducing the individual items from 35 to 12.\(^{137,138}\) The final 12 variables measured include: heart rate, mean arterial blood pressure, respiratory rate, temperature, Glasgow Coma Score, hematocrit, white blood cell count, serum potassium, serum sodium, serum creatinine, serum pH, and PaO\(_2\). The APACHE II is measured within the first 24 hours of ICU admission. The maximum severity score on the APACHE II is 71 with a score of 25 predicting a mortality rate of 50% and a score of 35 predicting a mortality rate of 80%.\(^{137,138}\) The APACHE II score was evaluated using an on-line calculator: [https://www.mdcalc.com/apache-ii-score].\(^{139}\)

**Sequential Organ Failure Assessment (SOFA)**

The SOFA score looks at the degree of organ dysfunction associated with sepsis. It has also been validated for organ dysfunction without signs of sepsis.\(^{137}\) The SOFA score looks at 6 domains: respiratory, coagulation, liver, central nervous system, cardiovascular, and renal. The SOFA is scored from 0-24 with 24 being the max severity.\(^{137,140}\) A SOFA score increase of 2 or more points within 24 hours of admission demonstrated an in-hospital mortality increase of
20.2% and an ICU length of stay increase to > 3 days.\textsuperscript{141} The SOFA score was evaluated using an on-line calculator: \url{https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score}.\textsuperscript{142}

**Statistical Analysis Review**

Quantitative analysis supports content validity of a physical functional measure. It is important to support that the instrument utilized is covering what it purports to measure.

**Classical Test Theory**

Classical Test Theory (CTT) is commonly known as true score theory. It assumes that every patient has a “true” score, and it would be achievable if there were no errors in the measurement. CTT indicates the observed test score ($X$) is equal to an individual’s true score ($T$) plus the measurement error ($E$), $X = T + E$. With CTT, the participant’s observed score on the whole measure is the focus. Test items need to be distinct enough from each other but similar and consistent with the same construct. A limitation to CTT is that the item parameters depend on the population studied, with difficulty of items dependent on the population tested. Random sampling from the specific population is required to obtain accurate estimates. A second limitation is that normal distribution of the scores is often assumed to be sufficient to allow for an interval scale. CTT lacks empirical evidence that the scale can be converted from ordinal to interval.\textsuperscript{143-145} CTT is widely utilized across various fields of study, including physical therapy, for test development and score analysis since the terminology and methodology are easily understood. Despite its frequent use, however, CTT is unable to estimate item difficulty and person ability separately. It is dependent on the population sample and the assumption of test item equivalence.
**Item Response Theory (IRT)**

IRT can be thought of as the theory of statistical estimation. The goal of IRT is to explain a connection between observed item responses and the underlying construct. IRT assumes that an individual’s performance on a test can be predicted based on the identification of latent traits or abilities, estimating ability scores, and subsequently using those scores to predict performance. With IRT, the test item is the focus. Test items need to be distinct from one another, but also similar and consistent with the same construct tested. With IRT, however, items are examined individually. One important feature within IRT, that is not evident with CTT, is the higher discrimination of items. For example, the distance between an individual’s current response and an item’s known severity has greater impact. Items that have greater discrimination provide more information to the individual’s status. IRT scores can adjust for difference in difficulty. A strength to IRT is that the sample does not need to come from a specific population. Non-random sampling can be completed, but within the sampling, the examinees need to span the range of item difficulties for accurate calibration of the item parameters.

**Rasch Analysis**

Rasch analysis identifies the strengths and weaknesses of a rating scale dissecting its structural and construct validity. Rasch analysis is used to determine the dimensionality of a scale, ensuring uni-dimensionality of the construct being measured. The Rasch model can transform ordinal, non-equal-interval scores into linear, equal-interval units. Unlike CTT, the measurement is not variable to the person taking the test. Rasch measurement, however, is not altered to fit the data like IRT. Rasch analysis further tests assumptions that ordinal-level scores approximate interval level scores by conversion of raw scores. Although a 1-parameter IRT may
look identical to a Rasch model, they are fundamentally different in how they alter the data to fit.\textsuperscript{56} They can, however be used simultaneously allowing for a complete evaluation of both the persons and items.

**Statistical Analysis for this Dissertation**

A recent study by Petrillo and colleagues\textsuperscript{146} indicated that IRT and Rasch analysis provided more diagnostic details to improve patient reported outcome scales. IRT and Rasch were not only able to target the areas of concern but were also able to flag potential causes, which assisted with correction of the scale. Another study by Hamel and colleagues\textsuperscript{147} compared CTT, IRT, and Rasch with missing data. IRT and Rasch were noted to be stronger and have less variability with missing data than CTT. IRT is identical to the dichotomous (1-parameter logistic) Rasch model, which is commonly used in the development and scoring of health-related outcome measures.\textsuperscript{148} This dissertation will be using IRT and Rasch Analysis as it offers benefits over CTT and will henceforth utilize the terminology “IRT Rasch analysis” to indicate both.

The purpose of Rasch analysis is transformation of ordinal, non-equal-interval scores into linear, equal interval units.\textsuperscript{56} Raw scores from latent traits (i.e., function) are converted into true “probabilistic” measures that have the same validity as other measuring tools (i.e., rulers, thermometers, or scales).\textsuperscript{56} Rasch analysis first creates linearity with equal measurement unit intervals; second, hierarchical arrangement of items demonstrates gradations of difficulty; third, independent objectivity indicates the tool can be used to measure other samples than from the one upon which it was built; fourth, identifies redundant items.
With IRT, the individual estimates are invariant and are not dependent on the group, item parameters transcend the population sample chosen, and predictive statements can be made as to an individual’s performance. IRT specializes in looking at an individual’s item-level response as opposed to the whole measure score. Items are ranked according to difficulty. Once completed, with an adequate pool of items, new items can be filled in to replace areas of difficulty as needed to allow for accurate functional capturing of patients across the continuum.

IRT Rasch Analysis will be used for 4 major reasons. First, to ensure that the tool is unidimensional, and only one construct is being measured: physical function. Second, item difficulty is ranked appropriately to allow for the full spectrum of function to be captured. Third, since person separation is identified, items are able to distinguish between levels of the construct. And finally, data can be converted to an equal-interval scale.

Summary

There have been 4 recent systematic reviews\textsuperscript{36,37,43,52} that look at outcome measures in the ICU. Vanpee et al.\textsuperscript{52} looked specifically at limb muscle strength assessment, but the other 3 systematic reviews looked at functional assessment. Each article stresses the need to select the appropriate measure to evaluate efficacy of services delivered, to identify change over time, and to have robust clinimetric properties. Measures developed for only one setting or patient population need to be extrapolated with caution. Patient alertness, sedation practices, delirium, severity of illness, and time are all factors that influence the completion of an outcome measure. Interventions aim to prevent the negative sequelae that are often associated with an ICU stay. An outcome measure needs to capture even small changes to allow for rapid identification and perhaps changes in interventions to address the concerns.
A systematic review by Tipping and colleagues\textsuperscript{32} examined outcome measures being used in PT clinical trials. The recommendations noted for specific PT interventions and patient safe handling in the ICU need to be interpreted with caution. The study revealed 19 outcome measures used across the 11 studies discussed. Only one outcome measure, used within these 11 studies, had clinimetric properties evaluated for use in the ICU. Synthesizing results and reporting recommendations was difficult in light of the varied outcome measure usage.

Another systematic review completed by Peterson and colleagues,\textsuperscript{36} identified 14 physical function measures that had some clinimetric properties related to the ICU. Gaps were noted in each of the measures utilized, thus limiting identification of one assessment tool that could serve as a gold standard. Identification of one measure that is comprehensive and sufficiently cohesive to evaluate the functional status of patients from dependent to independent across the spectrum of physical activity is lacking. Future studies need to be conducted with outcomes measures that are validated and reliable for the ICU setting.

**The Contribution of this Study**

This dissertation has addressed the gaps in the current physical function outcome measures by creating a comprehensive, cohesive physical function measure that is valid and reliable for use with patients in an ICU. The goal was to capture all levels of function from bed level to ambulatory, dependent to independent, and to limit the need for multiple outcomes measures to be used with a patient regardless of acuity and/or diagnosis. Selecting optimum tasks for evaluation to reduce redundancy and to provide a means for accurate, rapid assessment of functional decline was key to the functional measure design. Though all 14 of these known physical function measures have been previously tested in the ICU, many had not undergone rigorous clinimetric testing and therefore benefited from IRT Rasch Analysis.
Chapter 3: Methodology

Introduction to the Methodology Chapter

This chapter provides further detail into the theoretical framework and methodology used in this study. Details include research method, participants, research setting, enrollment process, and data collection, and analysis.

Research Method

This study is a single center study utilizing IRT Rasch Analysis to identify gaps in current physical function measures utilized in the ICU setting, provide an item difficulty ranking for each physical functional task, and create a comprehensive physical function measure to assess patients in the ICU. Participants were recruited using convenience sampling from cardiac, cardiothoracic, medical, and neurological ICU populations.

Specific Procedures Employed

Ethical Approval/Institutional Review Board

Institutional Review Board (IRB) approval was obtained from Our Lady of Lourdes Medical Center IRB Protocol Number 19-005/IRBNet Number 1410962-1. All protected health information was managed according to the HIPPA Guidelines.

Patients were coded numerically (i.e., 1, 2, 3…) as they joined the study. All key information retained for the study was aligned with the patient’s coded number allowing any further patient identifying information to be shredded. The only patient identifying information retained for study purposes was in accordance with the “limited data set” regulations set by HIPPA, which include age, ICU admission date, and ICU discharge date.
All patient data were initially recorded on a paper spreadsheet for ease of use during data collection on the ICU. The data were then copied to a computer spreadsheet. All paper data and signed consent forms remained in a locked cabinet and all computer data remained on a locked/pass-coded computer to which only the primary researcher had access. See Appendix D for the patient data spreadsheet.

Participants

This study was conducted at a large hospital in the Northeast United States. The institution is a teaching facility, which is a 325-bed destination hospital with one of the largest heart care and neurosurgical programs in the Delaware Valley. It is the only facility in southern New Jersey offering kidney, liver, and pancreas transplants. Other surgical specialties include vascular, general, bariatric, gynecologic, orthopedic, and urologic. There are two adult ICU’s. “ICU” is considered a medical/surgical/neurological unit, housing patients with a variety of diagnoses. “CCU” is a cardiac and cardiothoracic unit primarily including patients with cardiac related diagnoses. There is a total of 42 ICU beds in this facility. Participants were recruited from these (2) ICU settings as a means of convenience sampling.

To be eligible for this study, patients were admitted to the ICU or CCU as a direct transfer from an outside facility, the emergency room, or the operating room, and had an active “PT consult” order in place with appropriate activity orders (i.e., bed rest orders will not be considered for treatment). All patients who met the inclusion criteria were eligible to participate in the study. All patients who met the inclusion criteria were informed on the examination, assessment, and risks associated with the treatment session as they would for any physical therapy treatment session. Signed consent was obtained from those agreeing to participate as acknowledgement that subjects are fully aware of the study purpose and procedures.
Patients were able to decline to participate in the study if they chose. The patient was informed that should they choose to decline to participate in the study, they would still be evaluated by physical therapy in accordance with the physician order request and the standard-of-care but no data would be recorded from this patient except for standard medical charting.

**Recruitment**

A staff meeting was held for nursing staff from both ICU units and the physician intensivists to inform these personnel of the study purpose and methodology as well as to request support for the increased physical therapy time that would be needed with each patient. Awareness of the study helped to facilitate accurate nursing staff and physician response to general questions that patients had. Staff orientation also helped to assure patients that their medical team is supportive of the study and encourage patients to participate.

Two physical therapists associated with the study, each assigned to one of the two hospital ICU units served as screeners and data collectors for the study. These two therapists reviewed the medical record of patients referred to their respective ICU unit for physical therapy to determine patients that met the inclusion criteria and were medically stable to begin PT. To reduce the risk of bias and coercion, an occupational therapist (OT) assigned to the ICU’s recruited patients into the study. This OT informed the patient of the study and obtained signed consent. If the OT was not available for concurrent days, such as sickness or vacation, the two ICU physical therapists obtained the consent from patients in the ICU of the other physical therapist, and not from the ICU patients they were treating.
Research Personnel

The principal investigator, serving as one of the two physical therapists that collected data, has been a lead ICU physical therapist for more than 10 years. She has her Doctor of Physical Therapy (DPT) degree and is a Board Certified Cardiovascular and Pulmonary Specialist. The principal investigator was responsible for IRB submission, creation of the comprehensive physical function measure, recruitment, data collection, data analysis, and dissemination of results. The co-investigator, serving as the second data collector, has more than 15 years as an acute care therapist with the last 5 years specializing in cardiovascular and cardiothoracic ICU care. She has her Master of Physical Therapy degree (MPT) and is the lead therapist on the CCU. The co-investigator was also responsible for recruitment and data collection on study participants.

Patient Consent

Consent to participate in the study was obtained from the patient prior to patient assessment. The patient was informed by the study occupational or physical therapist that the patient was being asked to participate in a research study conducted by a PhD candidate from Nova Southeastern University. The patient was informed that the intent was to score them on 46 functional tasks as part of their treatment. The patient was also informed they qualified to participate in this study because they were admitted to the intensive care unit, their attending physician had referred them for physical therapy, and their participation in the study would include assessment and treatment as part of their normal course of treatment prescribed by their physician.
The patient was advised that their participation in this research study was strictly voluntary and that they had the right to withdraw their consent or discontinue participation in the study at any time without penalty. The patient was further informed that their current or future physical therapy or overall medical care at this facility would not be jeopardized if they chose not to participate in the study.

The ability to give consent was determined by having the patient answer four orientation questions: 1) knowledge of person, 2) knowledge of place, 3) knowledge of time, and 4) knowledge of event. The patient was also asked to explain, in general terms to the consenting therapist, the intent of the study and the patient’s understanding that they may refuse to participate at any time.

**Inclusion Criteria:**

Patients recruited for this study were newly admitted to the ICU or CCU, had active PT orders, and an appropriate activity order as defined previously. Further use of the term ICU will include CCU as well as ICU. Patients were 18 years or older and had the ability to give signed consent or had a family member/power of attorney (POA) present willing to give consent. Expected length of stay (LOS) in the ICU was greater than 48 hours as determined in discussion with the attending physician during morning rounds to limit those patients on the unit for observation status. The patient had a Richmond Agitation-Sedation Scale (RASS) score between -2 light sedation to +1 restless as determined by the ICU physical therapist during morning rounds with the physician.\textsuperscript{135}

A patient was included in the study once they were deemed hemodynamically stable by the medical team during regularly scheduled morning rounds. Vitals signs were noted upon
initial medical rounds to indicate that inclusion criteria had been met. Vital signs were again assessed prior to actual examination and start of data collection. Hemodynamic stability was blood pressure, heart rate, or pulse oximeter readings that had been stable for the last 4 hours with or without the use of pressor medications, mechanical ventilation, non-invasive ventilation (such as Bi-Pap or Vapotherm), or rate controlling medications. The 2018 PADIS guidelines for cardiovascular, respiratory, and neurological stability was utilized as criteria to initiate physical therapy unless otherwise indicated by the attending physician. These included: ability to open eyes to command, HR 60-130bpm, MAP 60-100mmHG, SBP 90-180mmHg, SPO2 > 88%, FiO2 < 0.6, and PEEP < 10. See Appendix D: Part B for charting.

Exclusion Criteria

Patients were excluded if they declined to provide signed consent, were hemodynamically unstable, or a family member/POA was required but was unavailable to provide consent to the study. The patient was excluded if they were transferred to the ICU from another in-patient hospital unit such as a telemetry or medical-surgical unit, indicating that the ICU was not their initial admission location from either the emergency room or the operating room. Patients that had a baseline functional status that was bedbound and/or non-ambulatory were also excluded from this study.

Sample Size Estimation

When utilizing IRT Rasch analysis a specific sample size is required to yield item separation as well as item reliability. For item calibration stability with plus/minus one-half logit at 99% confidence the sample size is 150. For item calibration stability “definitive or high
stakes” at 99% plus (items) confidence the sample size is 250. The proposed sample size for this study, based on these criteria, is 150 participants.

**Procedure**

Once patients met the requirements for inclusion into the study, the following demographic information was extrapolated from the patient’s chart by the primary researcher: age, gender, and ethnicity. Additional medical information was also obtained by the primary researcher such as admission diagnosis, ICU admission date, ICU length of stay (LOS), the APACHE II score (on admission and discharge from the ICU), and the SOFA score (on admission and discharge from the ICU). Refer to Appendix D for the data collection form.

**Instrument**

Table 3 lists 47 functional assessment tasks and 6 medical complexity characteristics that have been derived from the previously identified and currently utilized ICU physical function assessment measures. The final item under *medical status* served as a screening question and was not included in the analysis. These items have been set to a Likert scale for the Rasch model building by the principal investigator. All study subjects were evaluated with the 53 items on a comprehensive physical function evaluation by the two study physical therapists. The study subjects were not required to complete any other tasks besides those listed. Instructions for administration of the Comprehensive Physical Function Assessment (CPFA) tool can be found in Appendix E.
Table 3: The Comprehensive Physical Function Assessment (CPFA)

**Category I. Strength Testing Bilaterally**

<table>
<thead>
<tr>
<th>Test Description</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deltoid Right/Shoulder Abduction</strong>*</td>
<td>O</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Deltoid Left/Shoulder Abduction</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Shoulder Flexion Right</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Shoulder Flexion Left</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Biceps Right</strong>*</td>
<td>O</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Biceps Left</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Triceps Right</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Triceps Left</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Wrist Extensors Right</strong>*</td>
<td>O</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Wrist Extensors Left</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Iliopsoas Right/Hip flexion</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Iliopsoas Left/Hip Flexion</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Quadriceps Femoris Right/Knee extension</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Quadriceps Femoris Left/Knee Extension</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Tibialis Anterior Right/Dorsi-flexion</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Tibialis Anterior Left/Dorsi-flexion</strong>*</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td><strong>Grip Strength Right</strong></td>
<td>Unable</td>
<td>&lt; 20 kg</td>
<td>21-30kg</td>
<td>31-40kg</td>
<td>&gt;41kg</td>
</tr>
<tr>
<td><strong>Grip Strength Left</strong></td>
<td>Unable</td>
<td>&lt; 20 kg</td>
<td>21-30kg</td>
<td>31-40kg</td>
<td>&gt;41kg</td>
</tr>
<tr>
<td><strong>Completion of a Straight Leg Raise</strong></td>
<td>Unable</td>
<td>&lt; 15°</td>
<td>&gt; 15°&lt; 30°</td>
<td>30°</td>
<td>30° with 5 second hold</td>
</tr>
<tr>
<td><strong>Complete MRC-SS</strong>* (* indicates measure needed to create composite score)**</td>
<td>&lt; 10</td>
<td>11-25</td>
<td>26-36</td>
<td>37-48</td>
<td>49-60</td>
</tr>
</tbody>
</table>
## Category II. Bed Mobility

### Rolling Right

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Rolling Left

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Bridging

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Supine to Sit

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Sit to Supine

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

## Category III. Balance

### Static Seated Balance (unsupported)

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Dynamic Seated Balance (unsupported)

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Static Standing Balance (unsupported)

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Stand with feet together

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Stand on toes/heel raises

<table>
<thead>
<tr>
<th>1) Unable</th>
<th>2) Able to initiate but cannot clear heels from floor</th>
<th>3) Able to complete a heel raise but not able to hold</th>
<th>4) Able to complete heel raise but holds ≥5 seconds</th>
<th>5) Able to complete heel raise and holds ≥10 seconds</th>
</tr>
</thead>
</table>

### Semi-tandem stance (eyes open)

<table>
<thead>
<tr>
<th>1) Unable or assist of 2</th>
<th>2) Mod/Max Assist</th>
<th>3) CG/Min Assist</th>
<th>4) Supervision</th>
<th>5) Mod I / I</th>
</tr>
</thead>
</table>

### Tandem Stance with eyes closed
<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pick up a pen from the floor</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

**Category IV. Transfers**

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sit to Stand (hips at 90deg angle)</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stand to Sit</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bed to Chair/Transfer</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair to bed/transfer</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stand without use of arms (hips at 90deg angle)</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

**Category V. Gait**

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step Cadence</strong></td>
</tr>
<tr>
<td>1) Unable</td>
</tr>
<tr>
<td>2) 1-49 steps/min</td>
</tr>
<tr>
<td>3) 50-79 steps/min</td>
</tr>
<tr>
<td>4) 80-99 Steps/min</td>
</tr>
<tr>
<td>5) &gt; 100 steps/min</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ambulatory assistance without device</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ambulatory assist with device</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ability to walk backwards (4 Steps)</strong></td>
</tr>
<tr>
<td>1) Unable or assist of 2</td>
</tr>
<tr>
<td>2) Mod/Max Assist</td>
</tr>
<tr>
<td>3) CG/Min Assist</td>
</tr>
<tr>
<td>4) Supervision</td>
</tr>
<tr>
<td>5) Mod I / I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distance Walked</strong></td>
</tr>
<tr>
<td>1) 0-25 feet</td>
</tr>
<tr>
<td>2) 26-50 feet</td>
</tr>
<tr>
<td>3) 51-99 feet</td>
</tr>
<tr>
<td>4) 100-199 feet</td>
</tr>
<tr>
<td>5) ≥ 200 feet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time of Continuous Walk (included with above distance walked)</strong></td>
</tr>
<tr>
<td>1)</td>
</tr>
<tr>
<td>2)</td>
</tr>
<tr>
<td>3)</td>
</tr>
<tr>
<td>4)</td>
</tr>
<tr>
<td>5)</td>
</tr>
<tr>
<td>Category VI. Stairs</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Ascend/Descend Stairs</strong></td>
</tr>
<tr>
<td><strong>Stair Assistance</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section VII. Medical Status</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respiratory Function</strong></td>
<td><strong>1)</strong> Mechanical Ventilation</td>
<td><strong>2)</strong> NIV (BiPAP or High flow NC)</td>
<td><strong>3)</strong> ≥ 6 lpm or 48% FiO2</td>
<td><strong>4)</strong> &lt; 6 lpm or 48% FiO2</td>
<td><strong>5)</strong> Room Air</td>
</tr>
<tr>
<td><strong>Cough Ability</strong></td>
<td><strong>1)</strong> Absent</td>
<td><strong>2)</strong> Stimulated with suction</td>
<td><strong>3)</strong> Weak, needs to be suctioned</td>
<td><strong>4)</strong> Weak, can suction self</td>
<td><strong>5)</strong> Consistent volitional cough</td>
</tr>
<tr>
<td><strong>Mental Alertness/Command following</strong></td>
<td><strong>1)</strong> Unable to follow</td>
<td><strong>2)</strong> &lt; 25% commands</td>
<td><strong>3)</strong> 26-50% command</td>
<td><strong>4)</strong> 51-75% command</td>
<td><strong>5)</strong> ≥ 76% command</td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td><strong>1)</strong> 8-10</td>
<td><strong>2)</strong> 5-7</td>
<td><strong>3)</strong> 2-4</td>
<td><strong>4)</strong> 1</td>
<td><strong>5)</strong> 0</td>
</tr>
<tr>
<td><strong>Number of lines or tubes present</strong></td>
<td><strong>1)</strong> 9-10</td>
<td><strong>2)</strong> 7-8</td>
<td><strong>3)</strong> 4-6</td>
<td><strong>4)</strong> 1-3</td>
<td><strong>5)</strong> 0</td>
</tr>
<tr>
<td><strong>Is the patient on any continuous intravenous drips that cannot be disconnected for mobilization?</strong></td>
<td><strong>1)</strong> Yes</td>
<td><strong>2)</strong></td>
<td><strong>3)</strong></td>
<td><strong>4)</strong></td>
<td><strong>5)</strong> Yes</td>
</tr>
<tr>
<td><strong>Does the patient have a current contra-indication for OOB mobilization?</strong></td>
<td><strong>1)</strong> Yes</td>
<td><strong>2)</strong></td>
<td><strong>3)</strong></td>
<td><strong>4)</strong></td>
<td><strong>5)</strong> Yes</td>
</tr>
</tbody>
</table>
Data Collection

Two physical therapists conducted the data collection. Both physical therapists were the current lead therapist for their respective ICU unit. The physical therapists met to review the handbook created by the principal investigator, including all definitions of tasks and how tasks should be performed. There were no novel tasks included, however the handbook provided definitions specific to the study and details of each task including assistive device usage, time constraints, or special set-ups within the unit. Questions that arose were addressed during the review. (see Appendix E)

For inter-rater reliability, the two study therapists independently assessed 5 patients in the ICU, utilizing the comprehensive physical function assessment. Each assessment was conducted consecutively, dependent on patient physiological status, to promote consistent patient response to the assessment. The two therapists scored their individual evaluation form without conversing about their findings. The principal investigator then recorded the results from the two therapists and conducted the data analysis to establish inter-rater reliability.

The comprehensive physical functional assessment form was used on initial evaluation and upon discharge from the ICU. If a patient was discharged from the ICU to another inpatient hospital unit over the weekend and neither study therapist was working, discharge data were collected the following Monday morning by one of the study therapists on the respective unit where the patient was transferred.

The comprehensive physical function evaluation form was utilized in paper format to provide quicker access and collection of functional results. The patient demographics and clinical data form (Appendix D) was provided to record demographic and clinical data, and
information needed to complete APACHE II and SOFA scores. All data collected were entered into an Excel spreadsheet and uploaded into Winsteps and SPSS programs for statistical analysis.

Participants completed the comprehensive physical function exam in the order of strength assessment, bed mobility, seated and standing balance, transfers, gait, and then stairs as ordered in Table 3. If a patient was seated out of bed prior to the start of the physical therapy evaluation, the bed mobility portion was assessed last to increase efficiency and reduce fatigue for the patient in having to perform the transfers twice. Participants used an assistive device only if indicated in the instructions for the functional task assessed. Rest breaks were provided to participants as needed. Participants were requested to complete as many tasks as they could unless the treating therapist noted clinical compromise, or the patient refused to continue. (See Appendix E for further instructions) The research therapists utilized the same goniometer, a Marathon Adanac 3000 digital stopwatch, and a Jamar Hydraulic Hand Dynamometer, for data collection.

Data Analysis

Demographic and clinical data, including age, gender, ethnicity, diagnosis, disposition, severity of illness, length of stay, and time to PT consult and evaluation, were analyzed using descriptive statistics. These descriptive statistics included percentages, means, standard deviations, and minimum/maximum. ICU length of stay was calculated from day of admission to the day of physician order clearing the patient to discharge directly from the ICU or for transfer to another hospital unit. Hospital length of stay was calculated from the day of admission to the ICU to the day the patient was discharged from the hospital, including all hospital units. The discharge dispositions were tracked upon discharge from the hospital, either directly from the ICU or from another hospital unit. The discharge locations were defined as: 1) “Home”
including the option for visiting PT services, set up with outpatient PT services, or no services being required; 2) “Acute Rehab” indicating a high intensity rehab center transfer; 3) “Sub-Acute Rehab/Skilled Nursing Facility” indicating lower intensity rehab centers or need for skilled nursing long-term care; 4) “Long-Term Acute Care Hospital” transfer, most commonly referred to for chronic ventilation needs; and 5) “Expired” acknowledging those that demised in the hospital.

The two study therapists conducted inter-rater reliability of the comprehensive physical function exam to establish agreement. Inter-rater reliability was evaluated for both the whole score and the individual items/sections. Reliability was established with the kappa statistic (k), which provides the amount of agreement that would be expected but also accounts for the amount of agreement potentially due to chance.\textsuperscript{149} Kappa agreement is appropriate for usage on a categorical scale. The kappa statistic ranges from 0.00 to 1.00, the closer to 1.00 the higher the reliability and agreement.\textsuperscript{149} Landis and Koch,\textsuperscript{150} suggest the following criterion of value k: $< 0$ poor agreement, 0.0 – 0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, 0.81-1.00 almost perfect agreement. Inter-rater reliability was conducted using SPSS for overall agreement and sub-category agreement for the comprehensive physical function measure.

The researcher used IBM SPSS Statistics Software version 23 (IBM Corporation 1994, 2020, Chicago, IL) for the demographic and clinical data, and predictive validity. The Winsteps program 4.1.0 (Linacre, J. M. (2020) Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com) was used for the IRT Rasch analysis. Preliminary analysis allowed for individual activity task evaluation, ranking of task difficulty, and removal of duplicate tasks. IRT Rasch analysis included: item fit, hierarchy of items with roles for
duplicates and omissions, reliability, scale dimensionality, and differential item functioning. Rasch analysis also allowed for conversion of ordinal level data into equal-interval Rasch scale, which can be used for parametric statistical analysis.

**IRT Rasch Analysis- Fit**

Fit describes how well data conform to the Rasch model. Person and item fit were assessed with the following calculations: person infit MNSQ (mean-squared) and ZSTD (z-standard deviation); person outfit MNS and ZSTD; item infit MNSQ and ZSTD; and item outfit MNS and ZSTD. Outfit MNSQ is a chi-square calculation measuring levels of association with the outfit ZSTD providing a t-test level of probability determining if the MNSQ occurred by chance. Parameter-level mean-square fit statistics were interpreted as follows: < 0.5 logits is less productive for measurement but not degrading and may produce misleading reliabilities and separations; 0.5-1.5 logits is productive for measurement; 1.5-2.0 logits is unproductive for construction of measurement but not degrading; and > 2.0 logits distorts or degrades the measurement system. Mis-fitting persons or items can be removed from final inclusion.

For clinical observation, reasonable item mean-square ranges for infit and outfit are 0.5-1.7 logits. Outfit MNSQ, is known to be more sensitive to outliers and makes identification of fit issues easier. This study aimed to include outfit logits between 0.5-2.0 as they do not degrade the measurement system. Mis-fitting items or persons were examined for possible cause of misfit and potential removal of item or person.

**IRT Rasch Analysis- Hierarchy of Items and Scale Structure**

Ordering and spacing of items are important for the assessment of a measure’s quality. This hierarchy is best displayed with the use of a Wright Map. A Wright Map can identify trends
and item gaps, review overlapping items for removal, and identify missing areas of a functional assessment while examining both persons and items simultaneously.

The original measure of 53 items was reviewed according to the Wright Map in conjunction with item fit, for duplication of tasks and potential omission of tasks from the final measure. First, the hierarchy scale was set so that the hardest to complete tasks were at the top with the easiest to complete tasks at the bottom. Second, the scale was reviewed for any gaps in task completion, such as too much spacing noted between tasks that were completed. A gap of more than one logit may indicate missing traits. Third, items were identified for overlay or redundancy. Redundant items provide similar measurement information so one or more was omitted. Indications of an optimal instrument is when the mean value of the person matches with the mean value of the item measure.56

**IRT Rasch Analysis- Reliability**

Both person and item reliability were evaluated. Reliability values range from 0.0-1.0. The closer to 1.0, the more reliable the scale. A second area that supports reliability with IRT Rasch analysis is the “real” person separation and “real” item separation. Values can range from 0-infinity. Person separation index values of 1.50 or higher are acceptable, but an index > 3.0 is considered excellent. An item separation index value of 1.5 or higher is required for analyzing at the individual level and 2.5 is required for groups.56

**IRT Rasch Analysis- Scale Dimensionality**

Scale dimensionality is the ability of an item to measure the expected concept. This can also be thought of as item convergence within a scale and item discrimination across the scale. Dimensionality looks at each item and whether those items are measuring one underlying
dimension/construct/trait or several separate dimensions/constructs/traits. Within scale dimensionality is item discrimination. Item discrimination is a scale’s ability to discriminate or differentiate between individuals at different levels of function. A high discrimination is key to a functional tool. Being able to identify small functional differences is important. When looking at item difficulty and discrimination, if there is too large of a gap between functional items and their difficulty, the patient will have to improve their skill on a larger scale to make a change to their score. This also affects responsiveness.

Local independence assumes that the responses of one item will be independent of the response to the other tasks. This is an assumption that the traits being measured are not influenced by another trait. A tool that has broad item difficulty and discrimination will have good scale dimensionality and will reflect and capture functional differences between patients. Once the measure is determined to be uni-dimensional, local independence can be assumed and construct validity can be established.

Dimensionality was analyzed to identify the variance and covariance of the tasks to ensure the scale has one dimension. Uni-dimensionality of a scale is suggested when the Rasch dimension explains $\geq 40\%$ variance of the data, the first contrast of Rasch residual explains less than $5\%$ variance of the data, and the eigenvalue of the first contrast is $\leq 2.0$. $^{151-153}$

**IRT Rasch Analysis- Differential Item Functioning (DIF)**

DIF evaluates if a scale measures the same way for different groups of people. DIF compares the item characteristic curve for two different groups. If an item exhibits DIF it means the item displays a different characteristic across the two groups. This could indicate that the
item is “unfair” for a particular sub-group of patients. DIF will look at rank order of items between groups and take into account the separation of items.

For this study, DIF was evaluated for both gender (M/F) and diagnosis. A DIF is not expected to be large for gender due to varied diagnoses upon admission, however the potential for bias still exists. A DIF is expected to occur between the diagnostic groups because of their traits and characteristics. For example, the cardiovascular surgery group can be expected to have increased difficulty with certain functional tasks, such as standing or rolling, due to their sternal precautions in comparison to those with other diagnoses that do not have these limitations. The neurological and neurosurgical groups can be expected to have a larger DIF in comparison to the other diagnostic groups, regarding areas of strength and perhaps even overall mobility dependent on the extent of neurologic impairment. A DIF contrast of > 0.64 logits indicates moderate to large DIF. Any DIF contrast value close to this level was examined individually with a probability value cut-off of < 0.05. Any item which indicated a large DIF (contrast value ≥ 0.64 with probability cut-off of < 0.05) was evaluated for bias and possible omission from the final measure.

**IRT Rasch Analysis- Probability Curves**

An additional way to analyze effectiveness and function of a measure is in the use of probability curves. These curves evaluate the probability of a particular response category being selected. The vertical access of a probability curve represents the probability of responses, while the horizontal access represents the difference between a respondent’s measure and a specific item’s measure. The visual presentation of the probability curves include “hills.” Each hill represents a probability of occurrence between the item and a patient. Clear delineation and equal height of hills is optimal for probability.
Within the probability tables is a value known as the Andrich threshold. The Andrich threshold reveals how difficult it is to observe a category (not how difficult an item is to complete). If items are observed equally, the Andrich threshold increases with category value. The absence of this increase reveals a “disordering,” and indicates a rarely observed item. Probability curves were analyzed for the final version of the comprehensive physical function measure.

**IRT Rasch Analysis- Equal-Interval Scoring**

Using Winsteps, a linear transformation can be made in scoring of a measure converting values from ordinal to an equal-interval Rasch scale. Linear transformation does not alter the scale distribution, it simply converts the logit readings to an equal interval scale for the total score from 0-100. The final version of the comprehensive physical function measure has both raw and equal-interval scoring available.

**Predictive Validity**

Predictive validity of admission and discharge scores was established for this sample by dividing the participants into two groups: participants who were able to discharge home versus participants who were discharged to all other locations, i.e., rehab centers or long-term acute care hospitals. A Receiver Operating Characteristic (ROC) curve plot of sensitivity vs. 1-specificity was generated for each potential cut off score on the scale. The area under the curve (AUC) reflected the predictive ability with a higher AUC value indicating a stronger predictive ability. An optimal cut-off value, maximizing the sensitivity and specificity, was determined utilizing the Youden index (J). The positive predictive value (PV+) was calculated to estimate the probability that a patient who met the cut-off for discharge to home actually went home. The negative
predictive value (PV-) was calculated to estimate the probability that a patient was able to go home, despite falling below the cut-off threshold.

A perfect test would have both a 100% sensitivity and specificity.\textsuperscript{154} The AUC was interpreted as follows: 0.90-1.00 = excellent, 0.80-0.90 = good, 0.70-0.80 = fair, 0.60-0.70 = poor, and 0.50-0.60 = fail.\textsuperscript{155} The value of $J$ ranges from 0-1, with 1 being prefect ability and is calculated by $J = (\text{sensitivity} + \text{specificity}) - 1$.\textsuperscript{154,156} A LR+ > 5 and a LR- < 0.2 represent relatively important effects. A LR+ 2-5 and a LR- 0.2-0.5 may be important. Values close to 1.0 represent unimportant effects.\textsuperscript{149} The ROC curve analysis will be presented with tables and figures.

Likelihood ratios assist with confidence that the predictions are correct and are not dependent on prevalence. A high positive likelihood ratio (LR+) will indicate the increased confidence in predicting that patients who score above the cut-off value will be able to discharge to home. A high negative likelihood ratio (LR-) indicates an increased false negative occurrence. A measure with strong predictive ability will have a high LR+ and low LR-.\textsuperscript{149}

**Summary**

This study provides an in-depth examination of current intensive care unit (ICU) physical function assessment measures for the purpose of synthesizing these into one comprehensive measurement tool that addresses multiple areas of function. IRT Rasch Analysis was used to ensure that the tool is unidimensional, item difficulty is ranked, person separation is identified, and that data can be converted to an equal-interval scale.
Chapter 4: Results

Introductions to the Chapter

Chapter 4 details the results of this study for the aims addressed previously. *Aim 1*: Identify physical-function measures currently utilized in the ICU that have been psychometrically tested. *Aim 2*: Analyze all measure constructs to determine redundancies and appropriateness for use in the ICU setting according to IRT Rasch. *Aim 3*: Create a comprehensive, robust functional measurement tool for use with patients in the intensive care unit. Tables and figures are utilized to help organize the chapter.

Inter-rater Reliability-Pilot Study

Prior to initiation of the current study, a pilot study was conducted to establish inter-rater reliability of the assessment tool. Five patients in the ICU were randomly selected by the primary researcher for inclusion in that assessment. Table 4 lists inter-rater reliability patient demographic and clinical data.

Table 4: Inter-rater Patient Demographic and Clinical Data

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Race</th>
<th>Diagnosis</th>
<th>APACHE II Score</th>
<th>SOFA Score</th>
<th>Evaluation physical function measure total score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rater 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rater 2</td>
</tr>
<tr>
<td>1.</td>
<td>92</td>
<td>F</td>
<td>C</td>
<td>AF RVR/Resp Failure</td>
<td>13</td>
<td>2</td>
<td>173/270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>171/270</td>
</tr>
<tr>
<td>2.</td>
<td>81</td>
<td>M</td>
<td>B</td>
<td>CABG</td>
<td>8</td>
<td>0</td>
<td>221/270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>220/270</td>
</tr>
<tr>
<td>3.</td>
<td>27</td>
<td>F</td>
<td>A</td>
<td>Guillain-Barre Syndrome</td>
<td>18</td>
<td>7</td>
<td>80/270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80/270</td>
</tr>
<tr>
<td>4.</td>
<td>70</td>
<td>F</td>
<td>C</td>
<td>Robotic CABG/VDRF</td>
<td>8</td>
<td>0</td>
<td>185/270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>181/270</td>
</tr>
<tr>
<td>5.</td>
<td>66</td>
<td>M</td>
<td>C</td>
<td>CABG</td>
<td>5</td>
<td>0</td>
<td>226/270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>220/270</td>
</tr>
</tbody>
</table>

Kappa coefficient for the overall score was 0.942, p < 0.001. The results were “almost perfect agreement” for each sub-category with all values ranging > 0.81. Table 5 shows the specific kappa values for each total score as well as each sub-category.

Table 5: Inter-Rater Reliability Scores

<table>
<thead>
<tr>
<th>Inter-rater Reliability</th>
<th>Kappa Score (k)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.942</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Sub-Categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Strength</td>
<td>0.966</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Bed Mobility</td>
<td>0.834</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Balance</td>
<td>0.840</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Transfers</td>
<td>1.000</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Gait</td>
<td>1.000</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Advanced</td>
<td>0.879</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>- Medical Complexity</td>
<td>0.944</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Description of the Sample for Analysis of the CPFM

The two study researchers recruited subjects between April 2019 and February 2020. During this period, specifically on the days when the two study therapists were available, there were 788 new patients admitted to the ICU and referred for PT evaluation. Six hundred and twenty-eight patients were transferred to the ICU from other internal medical floors or from outside facilities excluding them from participation. A total of 160 patients met the inclusion criteria and were recruited to participate in the study. Three patients refused to participate for personal reasons, with 157 total patients enrolled in the study. All participants, or their designated POA, signed the consent forms. Seven of the 157 patients did not receive an examination at the time of discharge, as a study therapist was not available. The time from the physician order to completion of the PT evaluation was an average 1.15 days. A total of 150 patients had complete data sets for both admission and discharge data. See Figure 1 for the flow
chart of study recruitment. All participants attempted to complete each of the 53 tasks of the comprehensive physical function measure. If a participant could not complete a task due to medical concerns (i.e., unable to lay supine due to respiratory complications) or limited functional status, they were given a score of dependent/assist of two on that item.

Demographic and clinical data of the 157 study participants are listed in Table 6 including age, gender, ethnicity, diagnosis, disposition, severity of illness (APACHE II, SOFA), ICU and hospital LOS, and time to PT consult and evaluation.

Figure 1: Participant Recruitment
Table 6: Study Patient Demographic and Clinical Data

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (%)</td>
<td>Male (M)</td>
<td>51%</td>
</tr>
<tr>
<td></td>
<td>Female (F)</td>
<td>49%</td>
</tr>
<tr>
<td>Age ((\bar{x}) (SD), Min/Max)</td>
<td>Total Sample</td>
<td>65.9 (13.9), 18/94</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>65.4 (11.3), 39/91</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>66.4 (16.2), 18/94</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td>Caucasian/White</td>
<td>73.9%</td>
</tr>
<tr>
<td></td>
<td>Black/African American</td>
<td>17.2%</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>7.0%</td>
</tr>
<tr>
<td></td>
<td>Oriental/Asian</td>
<td>1.3%</td>
</tr>
<tr>
<td></td>
<td>Indian</td>
<td>0.6%</td>
</tr>
<tr>
<td>Diagnosis (%)</td>
<td>Neuro/Neurosurgical</td>
<td>40.1%</td>
</tr>
<tr>
<td></td>
<td>Cardio/Cardiosurgical</td>
<td>39.5%</td>
</tr>
<tr>
<td></td>
<td>Medical (i.e. sepsis, resp failure)</td>
<td>13.4%</td>
</tr>
<tr>
<td></td>
<td>General Surgical</td>
<td>7.0%</td>
</tr>
<tr>
<td>Discharge Disposition (%)</td>
<td>Home (No PT need or with Home/OP PT)</td>
<td>58.6%</td>
</tr>
<tr>
<td></td>
<td>Acute/High Intensity Rehab</td>
<td>25.5%</td>
</tr>
<tr>
<td></td>
<td>Sub-acute/Low Intensity Rehab</td>
<td>9.6%</td>
</tr>
<tr>
<td></td>
<td>Long-term Acute Care Hospital (LTACH)</td>
<td>3.2%</td>
</tr>
<tr>
<td></td>
<td>Hospice/Expired prior to DC</td>
<td>3.2%</td>
</tr>
<tr>
<td>Severity of Illness ((\bar{x}) (SD), Min/Max)</td>
<td>APACHE II Admit</td>
<td>11.76 (7.0), 0/36</td>
</tr>
<tr>
<td></td>
<td>APACHE II Discharge</td>
<td>7.56 (3.8), 0/21</td>
</tr>
<tr>
<td></td>
<td>SOFA Score Admit</td>
<td>3.8 (3.8), 0/18</td>
</tr>
<tr>
<td></td>
<td>SOFA Score Discharge</td>
<td>1.13 (1.9), 0/10</td>
</tr>
<tr>
<td>Time from ICU Admission to PT Consult Order ((\bar{x}) (SD))</td>
<td></td>
<td>1.55 days (3.6 days)</td>
</tr>
<tr>
<td>Time from PT order to PT evaluation ((\bar{x}) (SD))</td>
<td></td>
<td>1.15 days (0.76 days)</td>
</tr>
<tr>
<td>Length of Stay ((\bar{x}) (SD), Min/Max)</td>
<td>ICU Hospital</td>
<td>7.14 (7.8), 2/55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.8 (8.6), 2/56</td>
</tr>
</tbody>
</table>

KEY: N = population size, \(\bar{x}\) = mean, SD = standard deviation, min = minimum, max = maximum, neuro = neurological, cardio = cardiovascular, resp = respiratory, OP = outpatient, PT = physical therapy, DC = discharge, APACHE II = acute physiology and chronic health evaluation II, SOFA = sequential organ failure assessment, ICU = intensive care unit, LOS = length of stay
Rasch Analysis for the 53-Item Comprehensive Physical Function Measure

Item Fit

The 53 items for the 157 patients were initially analyzed for both admission and discharge values by the primary researcher. The admission data revealed that patients were functioning at a lower level, completing only the “easier” tasks on the scale, and were not able to complete many of the other “harder” tasks on the CPFM. The discharge data demonstrated an improved task distribution, noting that patients were able to complete more tasks including some of the harder tasks that were not completed on admission. This pattern would be expected with progression of function from admission to discharge. To increase the distribution of data, both admission and discharge scores for each patient were utilized for the Rasch analysis.

To limit other possible confounding factors, data were also analyzed and evaluated with the original CPFA Likert scale values for upper and lower extremity strength with a differential scale coding for weak (W) versus the strong (S) side. Grip strength was also analyzed and evaluated with the original scale versus a Likert scale coding for the patient’s age-predicted grip strength. The comparison Likert scale values for the age-predicted hand grip strength were utilized from the CPAx, which can be found in Appendix E. The decision was made by the primary researcher to continue with data analyses utilizing the original CPFA strength values. See Table 7 for Likert scale grip strength with corresponding values. Separate Wright maps for the admission and discharge data as well as the infit/outfit dimensionality of the strength comparisons are in Appendix F.
Table 7: Hand Grip Likert Scale Values

<table>
<thead>
<tr>
<th>Likert Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive</td>
<td>Unable</td>
<td>&lt; 20kg</td>
<td>21-30kg</td>
<td>31-40kg</td>
<td>&gt;41kg</td>
</tr>
<tr>
<td>Physical Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age-Predicted</td>
<td>&lt; 20%</td>
<td>20-39%</td>
<td>40-59%</td>
<td>60-79%</td>
<td>≥80%</td>
</tr>
<tr>
<td>Values for Hand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grip Strength</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The Wright map image of the 53 items in the comprehensive physical function measure (Figure 2 below) demonstrates the ability of the measure to capture a variety of mobility levels. Numerous redundant items were noted at the lower levels of function with few hard items available for higher functioning individuals. It is notable that tandem stance was indicated as the hardest item overall to complete. This sample had participants that were able to complete all the tasks.

Among all functional tasks, four items were indicated as “easiest” to complete. These items included bicep strength right (Rbicep), mental alertness (alertne), quadriceps strength right (Rquad), and deltoid strength right (Rdeltoid) with calculated logit (standard error) values of -3.13 (0.16), -2.98 (0.18), -2.93 (0.15), and -2.77 (0.13), respectively. The five items that were indicated as “hardest” to complete were tandem stance (tandem), number of stairs completed (stairs), stair assistance (stairas), picking up a pen from the floor (penfrom), and step cadence (stepcad) with calculated logit (standard error) values of 3.46 (0.08), 3.11 (0.08), 3.04 (0.08), 2.99 (0.08) and 2.99 (0.09), respectively.
Figure 2: Wright Map of the 53-Item Comprehensive Physical Function Measure

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>Person - MAP - Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>#</td>
</tr>
<tr>
<td>5</td>
<td>#</td>
</tr>
<tr>
<td>4</td>
<td>.##</td>
</tr>
<tr>
<td>3</td>
<td>#######</td>
</tr>
<tr>
<td>2</td>
<td>#</td>
</tr>
</tbody>
</table>

Key: tandem = tandem stance, penfrom = pen from the floor, stairas = stair assist, stairs = number of stairs, stepcad = step cadence, semi_ta = semi-tandem stance, standon = stand on toes, jump = ability to jump, ambwode = ambulate without a device, backwar = backwards walk, distanc = distance walked, ambwdev = ambulate with a device, bed_cha = bed to chair, chair_b = chair to bed, feettog = stand with feet together, standnoa = stand no arms, sit_sta = sit to stand, sit_sup = sit to supine, stand_s = stand to sit, Rgrip = grip right, standba = static standing balance, sup_sit = supine to sit, Lgrip = grip left, Lroll = roll left, Rroll = roll right, bridge = ability to perform bridge, dynseat = dynamic seated balance, ivgts = presence of intravenous drips, time = time walked, seatbal = static seated balance, SLR = straight leg raise, pain = presence of pain, lines = presence of lines, MRCSS = medical research council sum score, resp_fu = respiratory function, LDF = dorsiflexion left, LShldFl= shoulder flexion left, Lbicep = bicep left, LWristE = wrist extension left, Rhipfl = hip flexion right, Ldeltoid = deltoid left, Lquad = quadriceps left, Ltricep = tricep left, Rshldfl = shoulder flexion right, cough = presence of cough, Lhipfl = hip flexion left, Rdeltoi = deltoid right, RDF = dorsiflexion right, Rtricep = tricep right, RwristE = wrist extension right, Alertne = mental alertness, Rbicep = bicep right, , Rquad = quadriceps right
Five items had a response category Outfit MNSQ (ZTSD) value exceeding 2.0: dorsi-flexion right (RDF) with a logit 2.08 (2.07), grip strength right (Rgrip) with a logit 2.18 (9.90), respiratory function (resp_fu) with a logit 2.27 (4.35), pain (pain) with a logit 9.90 (9.91), and presence of intravenous drips (IVGtts) with a logit 9.90 (9.91). For a complete list of item order and outfit values, see Appendix G.

Hierarchy

The Wright map was reviewed for trends, item gaps, and overlapping of items. First, the overall trend for items is what one would expect in this sample. The hardest items (located towards the top of the Wright map) are higher level functioning tasks that require more skill, strength, and endurance. The easiest items (towards the bottom portion of the Wright map) are lower level tasks. Most of these tasks are simple manual muscle test categories.

Secondly, there are gaps noted that would indicate an area that a patient needs to overcome to achieve the next level. The first gap is noted between logits -1 to 0; this likely represents the transition from basic neuromuscular/bedbound function, and transitioning ability for out of bed trials. The second gap is noted at the top of the scale, > 3 logits. There is an uneven distribution overall with the center mean of patients being almost 2 logits away from the center mean of the items as noted by the value “M". (Figure 2)

Third, there is overlapping, and redundancy noted within the tasks themselves. There is an instance between logits 1 and 2, when 5 items (ambulate with a device, bed to chair, chair to bed, stand with feet together, and standing without arms) arrive at the same logit, which could indicate that these 5 tasks would likely assess a similar level of function. A second incident occurs between logits -3 and -2.
Reliability

The person separation and “real” person reliability index for the original comprehensive physical function measure (6.94, \( \alpha = 0.98 \)) is large. The item separation and “real” item reliability index for the original comprehensive physical function measure (17.55, \( \alpha = 1.00 \)) is also large and confirms scale hierarchy. The Cronbach Alpha (KR-20) person raw score "test" reliability = 0.98 with SEM = 5.73.

Scale Dimensionality

Principal component analysis of the comprehensive physical function measure showed a 39.4% raw variance in the items, suggesting the presence of greater than one construct measured. The first cluster of unexplained variances had an eigenvalue of 6.3 (2.4%) and included dorsiflexion right, lines, coughing, respiratory function, and pain. The second cluster of unexplained variances had an eigenvalue of 5.3 (2.0%) and again highlighted lines, coughing, respiratory function, and pain. The items are more indicative of medical status than function, highlighting a possible need for exclusion to maintain uni-dimensionality of the measure. The presence of dorsiflexion may be a result of clinical contrast between the cardiothoracic and neurological population. These items were reviewed for possible exclusion from the final measure. (See Figure 3)
Figure 3: 53-Item CPFM Scale Dimensionality

<table>
<thead>
<tr>
<th></th>
<th>Eigenvalue</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance</td>
<td>265.1610</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw variance explained by measures</td>
<td>212.1610</td>
<td>80.0%</td>
<td>81.0%</td>
</tr>
<tr>
<td>Raw variance explained by persons</td>
<td>107.7994</td>
<td>40.7%</td>
<td>41.2%</td>
</tr>
<tr>
<td>Raw variance explained by items</td>
<td>104.3616</td>
<td>39.4%</td>
<td>39.8%</td>
</tr>
<tr>
<td>Raw unexplained variance (total)</td>
<td>53.0000</td>
<td>20.0%</td>
<td>19.0%</td>
</tr>
</tbody>
</table>

| Unexplained variance in 1st contrast | 6.3561 | 2.4% | 12.0% |
| Unexplained variance in 2nd contrast | 5.2900 | 2.0% | 10.0% |
| Unexplained variance in 3rd contrast | 4.7731 | 1.8% | 9.0%  |
| Unexplained variance in 4th contrast | 3.8012 | 1.2% | 5.8%  |
| Unexplained variance in 5th contrast | 2.5518 | 1.0% | 4.8%  |

Key: Blue Outline: Contrast 1: A = lines, D = pain, G = respiratory function, O = dorsiflexion right, P = cough. Red outline: Contrast 2: A = lines, D = pain, G = respiratory function, P = cough
Differential Item Functioning (DIF)

DIF was evaluated across two patient subgroups: gender and diagnosis. DIF was identified by a probability < 0.05 with an effect size > 0.64. See Tables 8 and 9. There were 3 items noted with DIF when compared for male versus female; bilateral grip strength and respiratory function. For the strength items, it is known that females are generally weaker than males for grip strength.\textsuperscript{157} It is reasonable to presume that admission diagnosis most likely accounted for the strength differences noted and possibly the respiratory function since 53.2\% of females were admitted for a neurological/neurosurgical diagnosis, (as compared to 27.5\% of the males). The DIF across the diagnostic groups was large and were all evaluated for possible exclusion.

Table 8: Differential Item Functioning: Gender (Male vs Female)

<table>
<thead>
<tr>
<th>Item</th>
<th>Probability ((&lt; 0.05))</th>
<th>Effect Size ((&gt; 0.64))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grip Right</td>
<td>(&lt; 0.001)</td>
<td>-1.97</td>
</tr>
<tr>
<td>Grip Left</td>
<td>(&lt; 0.001)</td>
<td>-1.57</td>
</tr>
<tr>
<td>Respiratory Function</td>
<td>0.0070</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 9: Differential Item Functioning: Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Item</th>
<th>Probability ((&lt; 0.05))</th>
<th>Effect Size ((&gt; 0.64))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuro/NeuroSx Vs Cardio/CardioSx</td>
<td>Deltoid Left</td>
<td>0.0077</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>Shoulder Flexion Left</td>
<td>0.0188</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Bicep Flexion Left</td>
<td>0.0312</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Wrist Extension Left</td>
<td>0.0012</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td>Grip Strength Right</td>
<td>(&lt; 0.001)</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>Grip Strength Left</td>
<td>(&lt; 0.001)</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>MRC-SS</td>
<td>0.0028</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>Rolling Right</td>
<td>(&lt; 0.001)</td>
<td>-1.25</td>
</tr>
<tr>
<td></td>
<td>Rolling Left</td>
<td>(&lt; 0.001)</td>
<td>-1.38</td>
</tr>
<tr>
<td></td>
<td>Supine to Sit</td>
<td>(&lt; 0.001)</td>
<td>-1.48</td>
</tr>
<tr>
<td></td>
<td>Sit to Supine</td>
<td>(&lt; 0.001)</td>
<td>-1.63</td>
</tr>
<tr>
<td></td>
<td>Standing on Toes</td>
<td>0.0488</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td>Resp Function</td>
<td>(&lt; 0.001)</td>
<td>-3.55</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Item</td>
<td>Probability (≤ 0.05)</td>
<td>Effect Size (&gt; 0.64)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td>Coughing</td>
<td>0.0166</td>
<td>-1.56</td>
</tr>
<tr>
<td></td>
<td>Pain</td>
<td>&lt; 0.001</td>
<td>-2.84</td>
</tr>
<tr>
<td></td>
<td>Lines</td>
<td>&lt; 0.001</td>
<td>-2.14</td>
</tr>
<tr>
<td></td>
<td>Deltoid L</td>
<td>0.0036</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Quad Extension Right</td>
<td>0.0325</td>
<td>-1.12</td>
</tr>
<tr>
<td></td>
<td>Standing on Toes</td>
<td>0.0488</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td>Distance Ambulated</td>
<td>0.0057</td>
<td>-0.78</td>
</tr>
<tr>
<td></td>
<td>Jump</td>
<td>0.0411</td>
<td>-1.06</td>
</tr>
<tr>
<td>Neuro/NeuroSx Vs Medical</td>
<td>Dorsiflexion Left</td>
<td>0.0264</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>Grip Strength Left</td>
<td>0.0499</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>Resp Function</td>
<td>&lt; 0.001</td>
<td>-3.10</td>
</tr>
<tr>
<td></td>
<td>Lines</td>
<td>0.0241</td>
<td>-1.99</td>
</tr>
<tr>
<td>Cardio/CardioSx Vs Medical</td>
<td>MRC-SS</td>
<td>0.0213</td>
<td>-1.12</td>
</tr>
<tr>
<td></td>
<td>Rolling Right</td>
<td>0.0207</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>Rolling Left</td>
<td>0.0162</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>Supine to Sit</td>
<td>0.0026</td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td>Sit to Supine</td>
<td>0.0065</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>Distance Ambulated</td>
<td>0.0317</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>Pain</td>
<td>0.0024</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>Lines</td>
<td>0.0474</td>
<td>1.49</td>
</tr>
<tr>
<td>Medical Vs Gen Sx</td>
<td>Grip Right</td>
<td>0.0455</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Grip Left</td>
<td>0.0455</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**Key:** Neuro = Neurological, NeuroSx = Neuro-surgical, Cardio = Cardiovascular, CardioSx = Cardio-Thoracic Surgery, GenSx = General Surgery, MRC-SS = Medical Research Council Sum Score, Vs = versus, IVGtts = intravenous drips

**Item Deduction Process**

The original scale, as noted, had 53 items. Several of the 53 items were noted to have redundancy, misfit, and DIF. Based on these factors the decision was made to remove 38 items. Four items were removed for dimensionality: respiratory function, pain, coughing, and lines. One additional item, IV drips, was removed for misfit. The 16 individual strength items
including two grip items were excluded, as several demonstrated DIF across diagnostic
categories as well as two of those individual items were mis-fitting the model. Six of the items,
standing with feet together, semi-tandem stance, tandem stance, standing without the use of
arms, ambulation backwards, and step cadence, were removed for clinical relevance. The
remaining 9 items: sit to supine, dynamic seated balance, static standing balance, stair assistance,
ambulation with a device, stand to sit, bed to chair, distance walked, and straight leg raise, were
removed for redundancy.

The redundant items were looked at individually from a clinical perspective, evaluating
which item would be included. For example, ‘bed to chair’ and ‘chair to bed’ were duplicates at
\( \sim 1.5 \) logits. Clinically, bed to chair within the ICU is often easier due to surface height
difference, however, this may not always be the case for every ICU and every bed to chair
transfer. Therefore it was decided by the primary researcher that ‘chair to bed’ would be utilized
for the statistical scoring of the final CPFM measure, however, the final measure item would be
worded as ‘bed ↔ chair:’ (score transfer from lower surface to higher surface), as transferring
from the lower surface to a higher surface would provide increased clinical confidence in the
patient’s ability to perform transfers overall.

Several additional analyses were conducted to support the removal of individual
duplicates and use of different strength assessment combinations. A decision was ultimately
made to keep the MRC-SS, as it reduced both dimensionality and DIF. Mental
alertness/command following was supported by the results from fit, hierarchy, dimensionality,
and DIF analysis. It was also deemed appropriate to keep the task of rolling to the left and
rolling to the right despite their redundancy due to specific functional limitations seen within the
neurological patient population.
The scale was ultimately finalized with the following 15-items ranked from easiest to hardest: 1) Mental alertness/command following, 2) MRC-SS, 3) static seated balance, 4) the ability to perform a bridge, 5) rolling to the left, 6) rolling to the right, 7) performing supine to sit, 8) performing sit to stand, 9) transferring from chair to bed, 10) the ability to ambulate without an assistive device, 11) the time of continuous walk, 12) the ability to stand on toes, 13) the ability to pick up a pen from the floor, 14) the ability to ascend/descend stairs, and finally 15) the ability to jump. See Figure 4: Process summary.

Figure 4: Process Summary

Phase 1
- Identification of physical-function measures currently utilized in the ICU
- Extrapolation of all individual tasks listed in the physical function measures identified to create a pool of test items

Phase 2
- 53 items identified and set to a Likert scale for creation of the Comprehensive Physical Function Measure
- Inter-rater reliability for use of the Comprehensive Physical Function Measure

Phase 3
- The use of IRT Rasch analysis to analyze all measure constructs to determine redundancies and appropriateness

Phase 4
- 53 items, 38 items excluded
  - Items removed for Dimensionality (4 Items)
  - Items removed for Outfit (3 items)
  - Individual Strength and Grip (16 Items)
  - Items removed to maximize clinical relevance (6 items)
  - Items removed for redundancy (9 Items)

Phase 5
- The second run of IRT Rasch analysis to analyze all the final measure constructs to determine any further redundancies
- Instrument refinement to 15 items
Rasch Analysis for the 15-Item Comprehensive Physical Function Measure

With the use of IRT Rasch analysis, the 53-Item CPFM resulted in a finalized 15-Item CPFM (CPFM-15). The following section presents the IRT Rasch analysis for the final measure created.

Item Fit

The 15 items for the finalized comprehensive physical function measure included mental alertness/command following, MRC-SS, rolling right, rolling left, supine to sit, bridging, static seated balance, stand on toes, picking up a pen from the floor, sit to stand, chair to bed, ambulatory assist without a device, time of continuous walk, ascend/descend stairs, and ability to jump. The Wright map image of the 15 items in the comprehensive physical function measure (Figure 5) demonstrates the ability of the measure to capture a variety of mobility levels. There is a noted balance between the people and items with the curve matching in the middle. This sample of patients in the ICU is also noted to have participants capable of completing all tasks.

The “easiest” item for participants to complete was mental alertness with a logit (standard error) of 1.28 (0.50). The item that was indicated as “hardest” to complete was jumping, with a logit (standard error) of 0.85 (0.81). Even though there are still two redundant items, rolling left and rolling right, it was prudent to maintain rolling in both directions to allow for diagnostic differences, specifically in the neurological patient population. None of the items were found to have misfit. See Table 10.
Figure 5: Wright Map of the 15-Item Comprehensive Physical Function Measure

Wright Map of the 15-Item CPFM. The item hierarchy is noted from easiest tasks to hardest tasks (Green arrows). The distribution is normal with the center mean of patients matching the center mean of the items indicated as “M” (Red Circle). The yellow outline indicates patients who were at a lower level of function and unable to complete many of the tasks. The purple outline indicates patients who were able to perform all items on the scale.

Key to terms: Jump = ability to jump, Stairs = ability to complete stairs, PenFromF = ability to pick up a pen from the floor, StandOnT = ability to stand on toes, Time = time of continuous ambulation, AmbWODev = ambulation without a device, Chair_Be = Transfer chair to bed, Sit_Stan = transfer sit to stand, Sup_Sit = supine to sit, RollL = roll left, RollR = rolling right, Bridge = ability to complete a bridge in the bed, Seatbala = static seated balance, MRCSS = medical research council sum score, Mentalal = mental alertness.
Table 10: Outfit values of the 15-Item Comprehensive Physical Function Measure

<table>
<thead>
<tr>
<th>ENTRY NUMBER</th>
<th>TOTAL SCORE</th>
<th>TOTAL COUNT</th>
<th>ITEM MEASURE</th>
<th>MODEL S.E.</th>
<th>INFIT MNSQ</th>
<th>OUTFIT MNSQ</th>
<th>PTMEASUR-AL CORR. EXP.</th>
<th>EXACT MATCH OR%</th>
<th>EXP%</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1388</td>
<td>305</td>
<td>-4.51</td>
<td>.13</td>
<td>1.21</td>
<td>1.75</td>
<td>1.92</td>
<td>2.23</td>
<td>.64</td>
<td>.68</td>
</tr>
<tr>
<td>2</td>
<td>1076</td>
<td>305</td>
<td>-9.6</td>
<td>.10</td>
<td>1.08</td>
<td>.98</td>
<td>.95</td>
<td>-3.36</td>
<td>.84</td>
<td>.85</td>
</tr>
<tr>
<td>3</td>
<td>1079</td>
<td>305</td>
<td>-9.8</td>
<td>.10</td>
<td>1.14</td>
<td>1.61</td>
<td>1.61</td>
<td>.10</td>
<td>.84</td>
<td>.85</td>
</tr>
<tr>
<td>4</td>
<td>1181</td>
<td>305</td>
<td>-1.51</td>
<td>.10</td>
<td>1.14</td>
<td>1.19</td>
<td>.89</td>
<td>-3.37</td>
<td>.79</td>
<td>.79</td>
</tr>
<tr>
<td>5</td>
<td>973</td>
<td>305</td>
<td>-.09</td>
<td>.10</td>
<td>1.25</td>
<td>2.78</td>
<td>1.18</td>
<td>1.44</td>
<td>.84</td>
<td>.86</td>
</tr>
<tr>
<td>6</td>
<td>1483</td>
<td>305</td>
<td>-5.78</td>
<td>.21</td>
<td>1.72</td>
<td>3.44</td>
<td>1.28</td>
<td>.69</td>
<td>.43</td>
<td>.50</td>
</tr>
<tr>
<td>7</td>
<td>1235</td>
<td>305</td>
<td>-2.80</td>
<td>.12</td>
<td>-4.55</td>
<td>.57</td>
<td>.57</td>
<td>-3.01</td>
<td>.84</td>
<td>.79</td>
</tr>
<tr>
<td>8</td>
<td>730</td>
<td>305</td>
<td>2.45</td>
<td>.10</td>
<td>1.42</td>
<td>4.30</td>
<td>1.26</td>
<td>1.32</td>
<td>.83</td>
<td>.87</td>
</tr>
<tr>
<td>9</td>
<td>638</td>
<td>305</td>
<td>3.91</td>
<td>.10</td>
<td>.90</td>
<td>.98</td>
<td>.64</td>
<td>-1.11</td>
<td>.84</td>
<td>.84</td>
</tr>
<tr>
<td>10</td>
<td>939</td>
<td>305</td>
<td>.38</td>
<td>.10</td>
<td>.57</td>
<td>5.81</td>
<td>.51</td>
<td>4.27</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>11</td>
<td>915</td>
<td>305</td>
<td>.58</td>
<td>.10</td>
<td>.54</td>
<td>6.30</td>
<td>.47</td>
<td>4.42</td>
<td>.92</td>
<td>.87</td>
</tr>
<tr>
<td>12</td>
<td>861</td>
<td>305</td>
<td>1.15</td>
<td>.10</td>
<td>.81</td>
<td>2.21</td>
<td>.68</td>
<td>-2.35</td>
<td>.90</td>
<td>.88</td>
</tr>
<tr>
<td>13</td>
<td>827</td>
<td>305</td>
<td>1.44</td>
<td>.11</td>
<td>1.07</td>
<td>.81</td>
<td>1.63</td>
<td>.35</td>
<td>.87</td>
<td>.87</td>
</tr>
<tr>
<td>14</td>
<td>542</td>
<td>305</td>
<td>4.31</td>
<td>.12</td>
<td>1.27</td>
<td>2.46</td>
<td>.85</td>
<td>-.38</td>
<td>.79</td>
<td>.81</td>
</tr>
<tr>
<td>15</td>
<td>592</td>
<td>305</td>
<td>3.30</td>
<td>.10</td>
<td>.96</td>
<td>-2.80</td>
<td>.70</td>
<td>-3.72</td>
<td>.82</td>
<td>.82</td>
</tr>
<tr>
<td>MEAN</td>
<td>963.9</td>
<td>305.0</td>
<td>0.0</td>
<td>.11</td>
<td>1.05</td>
<td>.1</td>
<td>.93</td>
<td>-.7</td>
<td>78.9</td>
<td>68.9</td>
</tr>
<tr>
<td>P,SD</td>
<td>271.5</td>
<td>0.0</td>
<td>2.76</td>
<td>.03</td>
<td>.31</td>
<td>3.2</td>
<td>.37</td>
<td>1.9</td>
<td>7.5</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Hierarchy

The Wright map was reviewed for trends, item gaps, and overlapping of items. First, the overall trend for items is what one would expect with this sample. The hardest items (located towards the top of the Wright map) are understandably tasks that require a high level of skill, strength, and endurance, i.e., stairs and jumping. The easiest items (towards the bottom portion of the Wright map) are low-level tasks, i.e., mental alertness, and basic neuromuscular exam.

Secondly, there are gaps noted that would indicate an area where the measure may be missing items that could capture a patient within that level of function. The majority of gaps, however, are no more than 1 logit, which is considered acceptable for a clinical outcome measure. There is a gap of 2 logits between the MRC-SS and seated balance, however despite multiple runs of the data, none of the tasks could achieve the desired fit to minimize the gap. The
largest gap continues to be noted at the top of the scale above logit 4. There is a ceiling effect as some person data exceeds the highest item level (purple square area on Figure 4). There is an overall even distribution, with the center mean of patients matching the center mean of the items as noted by the curve distribution and the “M” indicators circled on Figure 4. As indicated previously, there were two tasks, rolling left and rolling right, that over-lapped, however, the decision was made to include these items.

**Reliability**

The person separation and “real” person reliability index for the finalized comprehensive physical function measure (5.13, $\alpha = 0.96$) were found to be large. The item separation and “real” item reliability index for the adjusted comprehensive physical function measure (21.52, $\alpha = 1.00$) are also large and confirm scale hierarchy. The Cronbach Alpha (KR-20) person raw score "test" reliability = 0.96 with SEM = 2.72.

**Scale Dimensionality**

Principal component analysis of the finalized comprehensive physical function measure showed 38.3% raw variance in the items, which indicates the possibility of more than one-dimension present. There was one cluster of unexplained variance that had significance at an eigenvalue of 3.6 (3.5%), which included rolling right, rolling left, and supine to sit. (See Figure 6)
Figure 6: 15-Item CPFM Scale Dimensionality

<table>
<thead>
<tr>
<th></th>
<th>Eigenvalue</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance in observations</td>
<td>102.2615</td>
<td>100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw variance explained by measures</td>
<td>87.2615</td>
<td>85.3%</td>
<td>85.4%</td>
</tr>
<tr>
<td>Raw variance explained by persons</td>
<td>48.0997</td>
<td>47.0%</td>
<td>47.1%</td>
</tr>
<tr>
<td>Raw Variance explained by items</td>
<td>39.1618</td>
<td>38.3%</td>
<td>38.3%</td>
</tr>
<tr>
<td>Unexplained variance (total)</td>
<td>15.0000</td>
<td>14.7%</td>
<td>14.6%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>3.6074</td>
<td>3.5%</td>
<td>24.0%</td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast</td>
<td>2.0823</td>
<td>2.0%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Unexplained variance in 3rd contrast</td>
<td>1.4832</td>
<td>1.5%</td>
<td>9.9%</td>
</tr>
<tr>
<td>Unexplained variance in 4th contrast</td>
<td>1.2032</td>
<td>1.2%</td>
<td>8.0%</td>
</tr>
<tr>
<td>Unexplained variance in 5th contrast</td>
<td>1.0711</td>
<td>1.0%</td>
<td>7.1%</td>
</tr>
</tbody>
</table>

STANDARDIZED RESIDUAL CONTRAST 1 PLOT

<table>
<thead>
<tr>
<th>-9</th>
<th>-7</th>
<th>-5</th>
<th>-3</th>
<th>-1</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Key: Blue Outline: Contrast 1: A = rolling left, B = rolling right, C = supine to sit
Differential Item Functioning (DIF)

DIF was evaluated across two patient subgroups: gender and diagnosis. DIF was identified by a probability < 0.05 with an effect size > 0.64. There were positive DIF values noted within each subgroup. See Tables 11 and 12 for items, probability, and effect sizes. The items related to DIF are most likely to be related to gender and diagnostic differences, with females representing a higher percentage in the neurological group of subjects with strength differences and the males representing a higher percentage for the cardiothoracic subject group. The diagnostic differences are minimal when compared to the original scale and will be addressed in the discussion chapter.

Table 11: Differential Item Functioning (15-Item): Gender (Male vs Female)

<table>
<thead>
<tr>
<th>Item</th>
<th>Probability (&lt; 0.05)</th>
<th>Effect Size (&gt;0.65)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRC-SS Supine to Sit</td>
<td>0.0053 &lt; 0.001</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.83</td>
</tr>
</tbody>
</table>

Table 12: Differential Item Functioning (15 Item): Diagnosis

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Item</th>
<th>Probability (&lt; 0.05)</th>
<th>Effect Size (&gt;0.64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuro/NeuroSx Vs Cardio/CardioSx</td>
<td>MRC-SS</td>
<td>&lt; 0.001</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>Rolling Right</td>
<td>&lt; 0.001</td>
<td>-1.65</td>
</tr>
<tr>
<td></td>
<td>Rolling Left</td>
<td>&lt; 0.001</td>
<td>-1.82</td>
</tr>
<tr>
<td></td>
<td>Supine to Sit</td>
<td>&lt; 0.001</td>
<td>-1.96</td>
</tr>
<tr>
<td></td>
<td>Standing on Toes</td>
<td>0.0001</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Ambulate WO Device</td>
<td>&lt; 0.001</td>
<td>1.12</td>
</tr>
<tr>
<td>Neuro/NeuroSx Vs Medical</td>
<td>Mental Alertness</td>
<td>0.0007</td>
<td>-1.87</td>
</tr>
<tr>
<td></td>
<td>Standing on Toes</td>
<td>0.0240</td>
<td>-0.80</td>
</tr>
<tr>
<td></td>
<td>Jump</td>
<td>0.0466</td>
<td>-0.84</td>
</tr>
<tr>
<td>Neuro/NeuroSx Vs GenSx</td>
<td>Pen from the floor</td>
<td>0.0385</td>
<td>0.83</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>Item</td>
<td>Probability (≤ 0.05)</td>
<td>Effect Size (&gt; 0.64)</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Cardio/CardioSxVs Medical</td>
<td>MRC-SS</td>
<td>&lt; 0.001</td>
<td>-1.91</td>
</tr>
<tr>
<td></td>
<td>Rolling Right</td>
<td>&lt; 0.001</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>Rolling Left</td>
<td>&lt; 0.001</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>Supine to Sit</td>
<td>&lt; 0.001</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>Mental Alertness</td>
<td>0.0002</td>
<td>-2.05</td>
</tr>
<tr>
<td></td>
<td>Stand on Toes</td>
<td>&lt; 0.001</td>
<td>-1.69</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>0.0175</td>
<td>-0.83</td>
</tr>
<tr>
<td></td>
<td>Stairs</td>
<td>0.0157</td>
<td>-0.90</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>MRC-SS</td>
<td>0.0011</td>
<td>-1.85</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>Rolling Right</td>
<td>0.0046</td>
<td>1.26</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>Rolling Left</td>
<td>0.0090</td>
<td>1.14</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>Supine to Sit</td>
<td>0.0065</td>
<td>1.17</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>Pen from the Floor</td>
<td>0.0390</td>
<td>0.84</td>
</tr>
<tr>
<td>Cardio/CardioSxVs GenSx</td>
<td>Ambulate WO Device</td>
<td>0.0326</td>
<td>-0.90</td>
</tr>
<tr>
<td>Medical Vs Gen Sx</td>
<td>Supine to Sit</td>
<td>0.0365</td>
<td>-0.98</td>
</tr>
<tr>
<td>Medical Vs Gen Sx</td>
<td>Pen from the Floor</td>
<td>0.0422</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Key: Neuro = Neurological, NeuroSx = Neuro-surgical, Cardio = Cardiovascular, CardioSx = Cardio-Thoracic Surgery, GenSx = General Surgery, WO= without, MRC-SS = Medical Research Council Sum Score.

Probability Curves

The first evaluation of probability was conducted via the use of observed percentages for each possible response as noted in Table 13. The Likert scale, 1-5 option responses, demonstrates similar observed percentages for each response. There is no demonstration of an outlying response in relations to the others. The Andrich threshold is also correctly ordered indicating stable responses for each value.
The second step in analyzing the probability of items, was performed through the probability curve. (See Figure 7) The first observation is the delineation of each Likert value’s hill. Each value, 1-2-3-4-5, includes a large hilltop. Value “2” does have the smallest hill and this correlates with the lower observed percentage seen in Table 13 at 14%. However, though the response on “2” is smaller in percentage and has a smaller “hill,” it remains clearly delineated from the other values and is not disordered per the Andrich threshold.

Figure 7: Probability Curve for the 15-Item Comprehensive Physical Function Measure
Each of the 15-items for the comprehensive physical function measure have been evaluated for individual probability and can be viewed in Appendix H. Two individual items were noted to have disordering according to the Andrich threshold. These items were bridging and the ability to pick up a pen from the floor. The observation percentages and the respective Andrich threshold values can be seen in Table 14. Probability indicates the chance of a specific response. For the ability to bridge, the probability of scoring a 2- “Max/Mod,” or 3- “Min/CG” response was a much lower percentage than the other values. For the ability to pick up a pen from the floor, the probability between the responses 2- “Max/Mod,” 3- “Min/CG,” and 4- “supervision” were poor in observed values and hard to delineate.

In the event the Andrich threshold is disordered, the next step is to look at the number of respondents for the lower percentages. In the case of bridging, there were only 2% of participants who rated a 2- and 8% who rated a 3- on the scale. For the ability to pick up a pen, the majority (53%) of participants were not able to complete this task, with only 9% of patients able to complete this task at an independent level. The disproportional distribution of these two item responses may warrant a revision on the item or the responses. Clinically, these items have an important role in strength and balance abilities, and so were retained in the assessment tool. However, since the individual means for the disordered values were obtained from limited participants, and the items do not have outfit, these items should be evaluated further in future studies and samples.

Table 14: Observed Score and Andrich Values for Bridging and Pen from the Floor Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Observed Values (%)</th>
<th>Andrich Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to Bridge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-</td>
<td>14%</td>
<td>1- None</td>
</tr>
<tr>
<td>2-</td>
<td>2%</td>
<td>2- -0.11</td>
</tr>
<tr>
<td>3-</td>
<td>8%</td>
<td>3- -1.89</td>
</tr>
<tr>
<td>4-</td>
<td>36%</td>
<td>4- -0.89</td>
</tr>
<tr>
<td>5-</td>
<td>41%</td>
<td>5- 2.84</td>
</tr>
</tbody>
</table>

111
<table>
<thead>
<tr>
<th>Item</th>
<th>Observed Values (%)</th>
<th>Andrich Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to Pick up a Pen from the Floor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-</td>
<td>53%</td>
<td>1- None</td>
</tr>
<tr>
<td>2-</td>
<td>11%</td>
<td>2- -1.51</td>
</tr>
<tr>
<td>3-</td>
<td>17%</td>
<td>3- -1.75</td>
</tr>
<tr>
<td>4-</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>5-</td>
<td>9%</td>
<td>4- 0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5- 2.56</td>
</tr>
</tbody>
</table>

Scoring

The raw versus interval scoring for the 15-item comprehensive physical function measure can be seen below in Table 15. The raw scoring ranges from 15-75 with equal-interval Rasch level scoring ranging from 0-100. The equal-interval Rasch level score can be used in the future with parametric testing. Table 16 demonstrates the CPFM-15. When running the admission and discharge data in two separate Rasch analyses, the CPFM-15 yielded 0.0% floor and 1.9% ceiling effects from the admission data and 0.0% floor and 3.3% ceiling effects from the discharge data.

Table 15: Scoring of the 15-Item Comprehensive Physical Function Measure

<table>
<thead>
<tr>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
<th>SCORE</th>
<th>MEASURE</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>.00E</td>
<td>9.91</td>
<td>36</td>
<td>45.73</td>
<td>2.17</td>
<td>57</td>
<td>64.01</td>
<td>2.19</td>
</tr>
<tr>
<td>16</td>
<td>7.63</td>
<td>6.17</td>
<td>37</td>
<td>46.63</td>
<td>2.15</td>
<td>58</td>
<td>64.94</td>
<td>2.21</td>
</tr>
<tr>
<td>17</td>
<td>13.60</td>
<td>5.11</td>
<td>38</td>
<td>47.52</td>
<td>2.14</td>
<td>59</td>
<td>65.91</td>
<td>2.25</td>
</tr>
<tr>
<td>18</td>
<td>18.06</td>
<td>4.50</td>
<td>39</td>
<td>48.40</td>
<td>2.13</td>
<td>60</td>
<td>66.90</td>
<td>2.28</td>
</tr>
<tr>
<td>19</td>
<td>21.57</td>
<td>4.03</td>
<td>40</td>
<td>49.28</td>
<td>2.12</td>
<td>61</td>
<td>67.93</td>
<td>2.33</td>
</tr>
<tr>
<td>20</td>
<td>24.43</td>
<td>3.67</td>
<td>41</td>
<td>50.15</td>
<td>2.12</td>
<td>62</td>
<td>69.00</td>
<td>2.38</td>
</tr>
<tr>
<td>21</td>
<td>26.84</td>
<td>3.39</td>
<td>42</td>
<td>51.01</td>
<td>2.11</td>
<td>63</td>
<td>70.12</td>
<td>2.43</td>
</tr>
<tr>
<td>22</td>
<td>28.91</td>
<td>3.16</td>
<td>43</td>
<td>51.87</td>
<td>2.11</td>
<td>64</td>
<td>71.30</td>
<td>2.50</td>
</tr>
<tr>
<td>23</td>
<td>30.72</td>
<td>2.97</td>
<td>44</td>
<td>52.73</td>
<td>2.11</td>
<td>65</td>
<td>72.55</td>
<td>2.57</td>
</tr>
<tr>
<td>24</td>
<td>32.34</td>
<td>2.82</td>
<td>45</td>
<td>53.59</td>
<td>2.10</td>
<td>66</td>
<td>73.87</td>
<td>2.66</td>
</tr>
<tr>
<td>25</td>
<td>33.81</td>
<td>2.70</td>
<td>46</td>
<td>54.45</td>
<td>2.10</td>
<td>67</td>
<td>75.28</td>
<td>2.75</td>
</tr>
<tr>
<td>26</td>
<td>35.17</td>
<td>2.60</td>
<td>47</td>
<td>55.30</td>
<td>2.10</td>
<td>68</td>
<td>76.82</td>
<td>2.87</td>
</tr>
<tr>
<td>27</td>
<td>36.44</td>
<td>2.52</td>
<td>48</td>
<td>56.15</td>
<td>2.10</td>
<td>69</td>
<td>78.50</td>
<td>3.02</td>
</tr>
<tr>
<td>28</td>
<td>37.64</td>
<td>2.46</td>
<td>49</td>
<td>57.01</td>
<td>2.10</td>
<td>70</td>
<td>80.37</td>
<td>3.21</td>
</tr>
<tr>
<td>29</td>
<td>38.78</td>
<td>2.40</td>
<td>50</td>
<td>57.86</td>
<td>2.10</td>
<td>71</td>
<td>82.51</td>
<td>3.45</td>
</tr>
<tr>
<td>30</td>
<td>39.88</td>
<td>2.35</td>
<td>51</td>
<td>58.72</td>
<td>2.10</td>
<td>72</td>
<td>85.05</td>
<td>3.80</td>
</tr>
<tr>
<td>31</td>
<td>40.92</td>
<td>2.31</td>
<td>52</td>
<td>59.57</td>
<td>2.11</td>
<td>73</td>
<td>88.25</td>
<td>4.38</td>
</tr>
<tr>
<td>32</td>
<td>41.94</td>
<td>2.27</td>
<td>53</td>
<td>60.44</td>
<td>2.12</td>
<td>74</td>
<td>92.92</td>
<td>5.71</td>
</tr>
<tr>
<td>33</td>
<td>42.92</td>
<td>2.24</td>
<td>54</td>
<td>61.31</td>
<td>2.13</td>
<td>75</td>
<td>100.00F</td>
<td>9.74</td>
</tr>
<tr>
<td>34</td>
<td>43.88</td>
<td>2.21</td>
<td>55</td>
<td>62.19</td>
<td>2.14</td>
<td>75</td>
<td>100.00F</td>
<td>9.74</td>
</tr>
<tr>
<td>35</td>
<td>44.81</td>
<td>2.19</td>
<td>56</td>
<td>63.09</td>
<td>2.16</td>
<td>75</td>
<td>100.00F</td>
<td>9.74</td>
</tr>
</tbody>
</table>
Table 16: The 15-Item Comprehensive Physical Function Measure

**Comprehensive Physical Function Measure**

**Section I. Medical Status**

1. **Mental Alertness/Command following**
   - 1) Unable to follow
   - 2) < 25% commands
   - 3) 26-50% command
   - 4) 51-75% command
   - 5) ≥ 76% command

**Section 2. Strength Testing**

(Composite Score of Bilateral Shoulder Flexion, Elbow Flexion, Wrist Extension, Hip Flexion, Knee Extension, Dorsi-Flexion [0-5-scale])

2. **Complete MRC-SS**
   - 1) < 10
   - 2) 11-25
   - 3) 26-36
   - 4) 37-48
   - 5) 49-60

**Section III. Bed Mobility**

3. **Rolling Right**
   - 1) Unable or assist of 2
   - 2) Mod/Max Assist
   - 3) CG/Min Assist
   - 4) Supervision
   - 5) Mod I / I

4. **Rolling Left**
   - 1) Unable or assist of 2
   - 2) Mod/Max Assist
   - 3) CG/Min Assist
   - 4) Supervision
   - 5) Mod I / I

5. **Supine to Sit**
   - 1) Unable or assist of 2
   - 2) Mod/Max Assist
   - 3) CG/Min Assist
   - 4) Supervision
   - 5) Mod I / I

6. **Bridging**
   - 1) Unable or assist of 2
   - 2) Mod/Max Assist
   - 3) CG/Min Assist
   - 4) Supervision
   - 5) Mod I / I

**Category IV. Balance**

7. **Static Seated Balance (unsupported)**
   - 1) Unable or assist of 2
   - 2) Mod/Max Assist
   - 3) CG/Min Assist
   - 4) Supervision
   - 5) Mod I / I

8. **Stand on toes/heel raises**
   - 1) Unable
   - 2) Able to initiate but cannot clear heels from floor
   - 3) Able to complete a heel raise but not able to hold
   - 4) Able to complete heel raise but holds ≥5 seconds
   - 5) Able to complete heel raise and holds ≥10 seconds
9. Pick up a pen from the floor

| 1) Unable or assist of 2 | 2) Mod/Max Assist | 3) CG/Min Assist | 4) Supervision | 5) Mod I / I |

**Category V. Transfers**

10. Sit to Stand (hips at 90° angle)

| 1) Unable or assist of 2 | 2) Mod/Max Assist | 3) CG/Min Assist | 4) Supervision | 5) Mod I / I |

11. Bed ↔ Chair (score transfer from lower surface to higher surface)

| 1) Unable or assist of 2 | 2) Mod/Max Assist | 3) CG/Min Assist | 4) Supervision | 5) Mod I / I |

**Category VI. Gait**

12. Ambulatory assistance without device

| 1) Unable or assist of 2 | 2) Mod/Max Assist | 3) CG/Min Assist | 4) Supervision | 5) Mod I / I |

13. Time of Continuous Walk

| 1) Unable | 2) < 2 minutes | 3) 2-4 minutes | 4) 4-5 minutes | 5) > 5 minutes |

**Category VII. Advanced**

14. Ascend/Descend Stairs

| 1) Unable | 2) 1-3 steps | 3) 4-6 Steps | 4) 7-11 steps | 5) ≥ 12 steps |

15. Ability to Jump

| 1) Unable | 2) Able to bend at knees to initiate jump | 3) Knee flexion and attempted push off with PF noted. | 4) Able to initiate jump where 1-foot clears floor | 5) Able to jump with both feet clearing floor at same time |

**Raw Score:** ______________________________

**Equal-Interval Score:** ____________________
Predictive Validity

A receiver operating characteristic (ROC) curve was utilized to determine predictive validity for ability to discharge to home. Table 17 summarizes the ROC statistics and the prediction performances based on the corresponding cut-off values. For admission data, the AUC was 0.773, p < 0.001 with 95% confidence interval 0.700-0.847. The optimal cut-off values for predicting home versus another location for ICU admission scores were 42 raw & 51 equal-interval, with a sensitivity of 71.7%, a specificity of 67.7%, a positive predictive value (PPV) of 75.9%, a negative predictive values (NPV) of 62.9%, a positive likelihood ratio (LR+) of 2.22, and a negative likelihood ratio (LR-) of 0.42. For discharge scores, the AUC was 0.919, p < 0.001 with 95% confidence interval 0.878-0.960. The optimal cut-off values for predicting home versus another location for ICU discharge scores were 54 raw & 61 equal-interval, with a sensitivity of 81.6%, a specificity of 82.0%, a PPV of 86.8%, a NPV of 75.8%, LR+ of 4.53, and LR- of 0.22. Figures 8 and 9 display the ROC curves generated from the admission and discharge data, respectively.

Table 17: Predictive Validity of the 15-Item Comprehensive Physical Function Measure

<table>
<thead>
<tr>
<th>Score On ICU Admission and Discharge</th>
<th>Suggested Cut off Score (points)</th>
<th>Area under the curve (AUC)</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission Score</td>
<td>≥42</td>
<td>0.773</td>
<td>&lt;0.001</td>
<td>0.700 - 0.847</td>
<td>71.7%</td>
<td>67.7%</td>
<td>2.22</td>
<td>0.42</td>
</tr>
<tr>
<td>Discharge Score</td>
<td>≥54</td>
<td>0.919</td>
<td>&lt;0.001</td>
<td>0.878 - 0.960</td>
<td>81.6%</td>
<td>82.0%</td>
<td>4.53</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Key: Sig = significance, LR + = likelihood ratio positive, LR - = likelihood ratio negative
Figure 8: ROC Curve Admission Data

Figure 9: ROC Curve Discharge Data
Summary of Results

A total 150 complete admission and discharge data sets were analyzed. The mean age of the sample was 65.9 years. The majority patient sample was white/Caucasian (73.9%) and male (51%). The mean LOS in the ICU was 7.14 days and 9.8 days in the hospital overall. The two primary diagnosis categories were Neuro/Neurosurgical (40.1%) and Cardio/Cardiosurgical (39.5%). Among all patients, 58.6% were able to be discharged home.

Rasch analysis was utilized for individual item evaluation, ranking of task difficulty, and removal of duplicate tasks. IRT Rasch analysis revealed the presence of two dimensions, five items that were misfit, 18 items for DIF, 6 for clinical relevance, and 9 items for redundancy. The dimension of construct outside of physical function, was identified as a medical complexity/medical status dimension. The items coughing, respiratory functional status, presence of pain, presence of lines, and presence of intravenous drips were identified as a second dimension. These items also misfit the model and were not included within the physical function measure developed. The item mental alertness was maintained as it did not fall within the second dimension and was not misfit or causing DIF. Mental alertness is related to physical function enabling a therapist to determine a patient’s ability to perform and function.

Of the 53 original items, 15 items were selected for the final CPFM-15 rating scale. The items were ranked according to difficulty and, when compared from a clinical perspective, found to be appropriate. There was no misfit. The reliability indexes were high confirming scale hierarchy. DIF was deemed non-significant and the probability curves were well delineated and ordered. The CPFM-15 was found to have predictive validity for discharge to home with both ICU admission and discharge scores, while ceiling and floor effects were minimal. The CPFM-15 was found to be valid and reliable for patients in the ICU setting.
Chapter 5: Discussion

Introduction to the Chapter

This section provides an in-depth discussion of the study results and outcomes of the three study aims. The purpose of this study was to identify physical-function measures currently utilized in the ICU that had previously been psychometrically tested, allowing for extrapolation of individual tasks utilized in those measures, so that a new valid, reliable, comprehensive physical function measure could be created. Clinical applications, recommendations, and limitations of this study are also presented.

Aim 1: Identify physical-function measures currently utilized in the ICU that have previously been psychometrically tested.

There appears to be a need for a robust standardized outcome measure to objectively capture a patient’s functional status, track effectiveness of treatment interventions, solidify veracity in research, and to promote efficacy in clinical practice. There are more than 100 functional outcome measures available, however, utilization of an outcome measure in the ICU setting is limited. In 2015, Parry et al identified 26 measures currently utilized in ICU clinical research. Six of the 26 measures identified were developed specifically for the ICU setting, with these six having had limited psychometric testing. A recent systematic review identified 14 current physical function measures with psychometric testing completed for patients within the ICU. These measures included the: (1) PFIT, (2) CPAx, (3) Perme, (4) SOMS, (5) IMS, (6) FSS-ICU, (7) ACIF, (8) DEMMI, (9) SPPB, (10) EFA, (11) FCS, (12) MRC-SS, (13) HHD, and (14) FAB.
While there are several physical function measures readily available, many provide an incomplete picture of physical function. Gaps were noted in each of the above assessment measures, limiting identification of one assessment tool that could serve as a gold standard.\textsuperscript{36} For example, the PFIT does not assess bed mobility, gait, balance, or stairs. The CPAx limits strength testing to grip strength and does not assess gait or stairs. The Perme does not evaluate steps but does evaluate patients from bed mobility to gait, while including some medical status and strength components. Scoring, however, stops at ‘minimum assistance’ so patients score the same for ‘minimum assistance,’ ‘contact guard,’ ‘supervision,’ and ‘independent’ levels of function. The Perme scale is not sensitive to change limiting its ability to capture progression and thus, creates a high ceiling effect.

The SOMS is a categorical scale that denotes milestones of activity (i.e., 2: passive range of motion, 3: can sit in a chair, 4: can ambulate), but does not quantify scores within those general values. The IMS is also a categorical scale indicating milestones of activity but does not allow for quantifying progression of individual functional tasks. The FSS-ICU does not evaluate strength, balance, or stairs. The ACIF evaluates patients across a broad spectrum similar to the Perme but is limited in its scoring capacity. Patients are rated either as ‘unable,’ ‘dependent,’ or ‘independent.’ The potential for capturing change within a short amount of time is questionable. The DEMMI captures the full spectrum of function from bed mobility to gait and stairs but does not assess strength. However, there are concerns regarding the DEMMI being used with a critically ill patient as it requires repetition of tasks, including 6 different balance tests.

The SPPB does not assess strength, bed mobility, or stairs and requires patients repeat each task twice. The EFA focuses heavily on the cognitive domain with concerns for a vegetative state; little focus is given to mobility. The FCS scores are similar to the SOMS with categorical
milestones but little room for progression and responsiveness. The FAB does not cover bed mobility or strength and focuses heavily on activities of daily living such as feeding, washing, dressing, toileting. The MRC-SS along with hand-held dynamometry only addresses strength.

Upon inspection of the 14 currently available ICU measures, there are apparent deficits in each related to their ability to capture the spectrum of function from dependent to independent. This is, perhaps, why it is often necessary to utilize multiple outcome measures simultaneously. Dissection of the 14 ICU functional measures revealed 53 items/tasks that could be extrapolated to cover the entire spectrum of a functional outcome measure.

**Aim 2:** Analyze all measure constructs to determine redundancies and appropriateness for use in the ICU setting according to IRT Rasch analysis.

To support the structure and possibly supplement the 53 items that were extrapolated from the 14 current ICU measures, further review was conducted on the two ICU measures, the DEMMI\(^{115,116}\) and the PFIT,\(^{54}\) that had also undergone Rasch analysis during their development. The DEMMI started with a focus group of academics, clinicians, and patient interviews seeking consensus on item choice. Ninety-seven mobility items were generated from the focus groups. The DEMMI utilized many outcome measures for their pool of items including the Barthel Index, Timed-Up and Go (TUG), Katz ADL, and several other tools used in both the hospital and outpatient centers, further adding 75 items. One hundred twenty-one items were removed for item duplication, redundancy, and application of the inclusion criteria by two independent assessors. The final 51 tasks were then subjected to a two-week pilot test with 15 participants, on a general hospital ward for patients ≥ 65 years of age. Nine items were removed for their “practical limitations” by two independent assessors, leaving 42 items to undergo Rasch analysis.
The original pool of items for the DEMMI were extracted from measures used in various settings, i.e., inpatient hospital, community-dwelling, or outpatient settings. Items that were initially tested and omitted from the final DEMMI included: standing on one leg, toe and heel walking, number of times in and out of bed in 10 seconds, stepping over a box, hopping, 360 degree turn, and carrying a glass of water while walking. While most of these items have merit in specific settings, they do pose challenges in the ICU, i.e., the 360-degree turn with central or peripheral lines attached. All timed items (i.e., the timed up and go, number of times in/out of bed in 10 seconds) were removed from consideration for the CPFM as these also posed challenges to patients in the ICU setting. Ability to complete timed items is often not a limitation of the patient, but rather an issue in management of the lines and tubes associated within the ICU setting.

There were four items omitted from the final DEMMI that were included in the initial 53-item CPFM for repeat analysis: transferring from bed to chair, sitting to lying, semi-tandem stance, and steps. The ‘sit to supine’ and ‘transfer bed to chair’ tasks were part of the ACIF with the ‘semi-tandem stance’ being extrapolated from the SPPB. The ACIF and FAB also assess ‘steps.’ The DEMMI cited the reason for removal of bed to chair and steps as equipment needed, but not available for use during the assessment. However, the ICU unit used in the study always had the equipment available, including a bed for the patient and a chair at bedside. The facility utilized for this study also had steps on the floor and stools in the rooms due to the height of the beds. It was, therefore, deemed necessary to re-test these four items through IRT Rasch analysis within this specific ICU setting. The item for ‘steps’ was the only item of the four retained in the final CPFM-15.
The testing for the PFIT also supported the framework for the CPFM study. The PFIT was created specifically for use in the ICU through a pilot test and face validity. In 2013, Denehy et al. used the original PFIT created by Skinner et al. which consisted of 5 items (assistance to stand, step cadence, shoulder flexion strength, knee extension strength, and bilateral shoulder lifts per minute), and created the PFIT scored(s). Rasch analysis was utilized to analyze the 5 original items for dimensionality and misfit and then used to convert the total score from an ordinal scale to an equal interval. The PFIT(s) was finalized to four items with the removal of one item, ‘bilateral shoulder lifts per minute,’ due to misfit. The item removed from the original PFIT, ‘bilateral shoulder lifts per minute,’ was not included with the items chosen for the initial CPFM analysis, as it was already deemed misfit by Denehy et al. The only item retained within the final CPFM-15 from the PFIT(s) was ‘sit to stand.’ Review of the DEMMI and PFIT-PFITs revealed no further items needed to be added to the initial 53-Item CPFM.

The initial 53-Item CPFM was noted to have misfit, item gaps, redundant items, and DIF. The trend of the items would be what one would expect clinically with the “easiest” items to complete (mental alertness and strength testing) appearing at the bottom of the scale, while the “hardest” items (tandem stance, stairs, and step cadence) appear towards the top of the scale. There were very few gaps noted along the scale between the items, lending credence to a high discrimination of the tool to capture small functional difference. Too large of a gap would create a potential issue of insensitivity for the tool to capture the specific patient functional level.

The individual muscle strength assessment was noted to have the most redundancy as well as misfit and DIF. Several combinations of strength assessment were analyzed for improved fit, reduction of redundancy, and reduced DIF. However, despite the varied combinations, the misfit remained for most individual items. The decision was made, therefore, to maintain the
complete MRC-SS instead of assessment of individual muscle groups. Having the MRC-SS built into the CPFM-15’s structure limits the need for a secondary test, supporting the goal of developing an objective, efficient assessment tool. The use of the MRC-SS was further supported by Parry et al.,\textsuperscript{158} that recommended regular screening for muscle weakness using the MRC-SS when evaluating physical function in the ICU. Keeping the MRC-SS also allowed for complete assessment and early identification of ICU-acquired weakness (ICUAW).\textsuperscript{17,18} The MRC-SS is the only current way to clinically identify ICUAW without the use of Electromyography (EMG) or Nerve Conduction Study (NCS).\textsuperscript{17,18} The current accepted cut-off score for presence of ICUAW according to the MRC-SS is $< 48$.\textsuperscript{159}

The initial 53-Item CPFM was found to be reliable indicating that it will consistently measure items and patients the same way each time used, no matter what the patient’s functional level. The person sample was large enough to confirm the item difficulty hierarchy of the instrument, which supports the construct validity. Within the reliability of this tool, it is an important reminder that within IRT Rasch analysis there is person and item reliability, but also an overall reliability score reported with the use of Cronbach Alpha (KR-20). For those unfamiliar with Rasch analysis, the KR-20 is recognizable as reported within classical test theory. Within Rasch analysis, however, Cronbach Alpha is resulted from non-linear, raw data and always exceeds the maximum reliability possible, thus leaving its results corrupted.\textsuperscript{56} More statistical weight should be placed on the person and item reliability scores.\textsuperscript{56}

Within Rasch there is “real” reliability and “model” reliability. The “real” reliability values indicate the lower limit of the instrument’s consistency and is recommended for use in medicine as it is a more conservative estimate.\textsuperscript{160} In addition to the reliability value, however, Rasch analysis also provides a “separation” value.\textsuperscript{160} This is a “signal to noise” ratio (square
root value of the ratio between the true person or item variance and the error variance within the data). This indicates how well a set of items can differentiate different respondents. Item separation is used to verify item hierarchy. Separation values range from 0-infinity, so there is no ceiling.\textsuperscript{160} A higher value is better. Low ‘person’ separation (< 2) with a low ‘person’ reliability (< 0.8) implies that the instrument may not be sensitive enough to distinguish between high and low performers, more items may be needed.\textsuperscript{160} On the other hand, Low ‘item’ separation (< 3) with low ‘item’ reliability (< 0.9), implies that the person sample is not large enough to confirm the item difficulty hierarchy of the instrument and construct validity is affected.\textsuperscript{160} The CPFM-15 has high person and item reliability.

The scale dimensionality suggests the presence of greater than one construct. The clusters identified were more indicative of medical status than function. The dimensionality differences are suspected to be closely tied to the DIF, which was noted across the diagnostic groups. The DIF across diagnostic groups was large, but not completely unexpected. The DIF noted in the strength components were noted primarily in comparison between the neurological/neurosurgical group. This can be explained by the impairments that often accompany neurological injury versus a general weakness from surgery or general bedrest.

The second largest area of DIF was noted in a comparison between the cardio/cardiosurgical group and the other diagnostic groups. Bed mobility: rolling left, rolling right, and supine-to-sit, was notably harder for the cardio/cardiosurgical diagnostic group. This may be explained by the presence of sternal precautions among the majority of the cardio/cardiosurgical patients. With sternal precautions, patients are generally not permitted to use their hands to roll or perform supine to sit. Since the DIF findings do not necessarily indicate
bias, a decision was made to maintain these items as they can support the ability of the assessment tool to differentiate skills between diagnostic groups.

A further look into the differences noted between the diagnostic groups, revealed a DIF for lines/tubes, pain, coughing and respiratory function. The cardio/cardiosurgical group had a noticeable DIF when compared with the neuro/neurosurgical (lines/tubes, pain, cough, respiratory function) and the medical group (pain, lines). A similar DIF for lines was noted between the general surgical group when compared with the neuro/neurosurgical group. Clinically, post-surgical patients often require more standard intravenous medications and lines/tubes then would a patient, for example, with a cerebrovascular accident.

The CPAx\textsuperscript{53} and Perme\textsuperscript{55} were two outcome measures that utilized the medical status/medical complexity items noted above. The CPAx evaluated respiratory function and coughing similar to this study for the CPFM, with both using a Likert scale. (See Table 18 below). The CPAx used a scale from 0-5, this study utilized a scale 1-5. Both measures address a steady progression from dependent status to independent, but in different increments. The Perme addresses mental alertness, respiratory status, pain, presence of lines/tubes, and presence of IV drips. The Perme, however, uses a dichotomous format. For example, “does the patient have pain” 0 = yes, 1 = no. The benefit to use of a Likert scale, as done with creation of the CPFM, is in the “degree” of an answer. For example, how much pain, not just the presence of pain.

Table 18: CPFM versus CPAx Respiratory Function and Coughing Likert Scale

<table>
<thead>
<tr>
<th>Likert Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory Function</td>
<td>CPAx</td>
<td>Mechanical Ventilation</td>
<td>Non-invasive Ventilation (BiPap or high flow NC)</td>
<td>≥ 6 Liters or 485 FiO2</td>
<td>&lt; 6 Liters or 48% FiO2</td>
<td>Room Air</td>
</tr>
</tbody>
</table>
### Likert Score

<table>
<thead>
<tr>
<th>Likert Score</th>
<th>CPAx</th>
<th>Coughing</th>
<th>CPFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Complete ventilatory dependence</td>
<td>Absent Cough</td>
<td>Absent cough</td>
</tr>
<tr>
<td>1</td>
<td>Ventilatory dependent with spontaneous breaths</td>
<td>Stimulated with suction</td>
<td>Cough stimulated with deep suction</td>
</tr>
<tr>
<td>2</td>
<td>Non-Invasive Ventilatory support (BiPap, or high flow NC)</td>
<td>Weak needs to be suctioned</td>
<td>Weak cough, sometimes able to clear</td>
</tr>
<tr>
<td>3</td>
<td>Intermittent need for high flow O2 (&gt; 15L)</td>
<td>Weak can suction self</td>
<td>Effective cough with airway clearance technique</td>
</tr>
<tr>
<td>4</td>
<td>Oxygen Therapy (&lt; 15L)</td>
<td>Consistent volitional cough</td>
<td>Consistent volitional cough</td>
</tr>
<tr>
<td>5</td>
<td>No oxygen Therapy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Key:
- CPFM: original comprehensive physical function measure
- CPAx: the Chelsea critical care physical assessment tool
- BiPAP: Biphasic positive airway pressure
- NC: nasal cannula
- FiO₂: fraction of inspired oxygen

In comparison to the above outcome measures, the CPFM was able to compare number of lines/tubes present, amount of pain reported, severity of respiratory status, and the number of IV drips currently in use since it used a Likert scale to differentiate values. Upon reviewing the Wight Map, Figure 2, it can be noted, that despite the differentiation of values for severity of medical complexity, the items for coughing, pain, respiratory status, and lines/tubes fall below the below the overall mean (M). The only item above the overall mean was IV drips, although it was noted to be redundant with other items. With the use of the Likert scale within this sample, DIF was noted for the medical status/medical complexity items. When these results were reviewed in combination with misfit and the concern for dimensionality, the decision was solidified supporting removal of the medical complexity items.

There were six items removed from the original 53-item CPFM due to redundancy, as well as clinical concern regarding subject task performance. These items included: standing with feet together, semi-tandem stance, tandem stance, standing without the use of arms, ambulation backwards, and step cadence. The balance related items (standing with feet together, semi-
tandem stance, and tandem stance) were difficult for patients to complete properly given the presence of central groin lines, lower extremity swelling, lower extremity surgical sites, or lower extremity wounds that made these positions difficult to obtain and maintain. The ability to ambulate backwards was also removed due to fear and apprehension noted by the patients. Given the redundancy of these tasks, it was determined that their removal would not limit the ability to assess functional performance.

The item, ‘standing without the use of arms,’ was removed as it only applied to a specific population (i.e., the patient post median sternotomy). The item for ‘sit to stand’ was retained within the final CPFM-15 and can be assessed with or without arms as needed to adhere to given precautions for specific populations. The final item removed was step cadence. Patients in the current study reported an ease with ambulation over time as compared to step cadence. The decision was made, therefore, to keep the item, time of continuous walk rather than step cadence for the endurance assessment.

**Aim 3: Create a comprehensive, robust functional measurement tool for use with patients in the intensive care unit.**

With the use of IRT Rasch Analysis, the 53-item CPFM was ultimately finalized to include 15-items. The CPFM-15 was found to have stable and appropriate item fit and hierarchy, attesting to construct validity, excellent reliability, unidimensional, no significant DIF, and good predictive validity of discharge to home. The DIF noted within the diagnostic groups is consistent with the neurological group of subjects having increased difficulty with strength testing and the cardio/cardiosurgical group having increased difficulty with bed mobility tasks. It was determined that these items did not exhibit another dimension. The DIF findings do not necessarily indicate bias, but rather that more than one item may measure differently for a
different subgroup of patients. This does not necessarily confound the results, but rather might indicate the ability of the assessment tool to differentiate skills between diagnostic groups. Patients with differing functional impairments due to their diagnosis (i.e., limitations due to sternal precautions versus a patient who has had a cerebrovascular accident) should not be expected to respond the same to individual tasks. The final 15-items are ranked according to difficulty and can capture patients across the spectrum of function from dependent to independent.

With regard to hierarchy, there were gaps noted that would indicate an area where the measure may be missing item(s) that could capture a patient within that level of function. The largest gap of approximately 2 logits was noted between the items MRC-SS and seated balance. With the use of Rasch, multiple runs of the data were performed in an attempt to close this gap, but the desired fit could not be achieved. A benefit to utilizing a Likert scale in scoring is that it assists with explaining where apparent gaps may not truly be gaps. If Figure 10 is compared with the previous Figure 5, it can be seen how the Likert scale indicated scoring of patients within those gaps. Between the items for the MRC-SS and seated balance, are the patients who scored lower on the items for mental alertness, MRC-SS, seated balance, and even rolling. The large gap noted above jumping is also explained by the limited number of patients able to complete higher-level tasks at an independent (5 on Likert scale) level. The ceiling effect was present, but low at 1.9% on admission and 3.3% on discharge.
purple square indicates the large gap at the top of the scale for more independent completion of hard tasks. Yellow square indicates the gap between MRCSS and seated balance that is filled with patients who completed lower level tasks.

Key: Standon: stand on toes, penfrom: pen from floor, ambwodev: ambulate without a device, sit_sta: sit to stand, chair_b: chair to bed, rollL: rolling to left, rollR: rolling to right, sup_sit: supine to sit, seatbal: seated balance, MRCSS: medical research council sum score, Mentala: mental alertness
The CPFM-15 was analyzed for effectiveness and function using probability curves. The curves evaluate the probability of a particular response category selected. The visual presentation was evaluated using the vertical access for the probability of responses, the horizontal access as the difference between a respondent’s measure and a specific item’s measure. Clear delineation of the hills is optimal for probability. (Refer back to Figure 7) In addition to the probability curves is a value known as the Andrich threshold. This reveals how difficult it is to observe a category (not to be confused with how item difficulty). If items are observed equally, the Andrich threshold will increase with the category value. The absence of this increase is known as disordering. The CPFM-15 is well ordered.

Two individual items noted to have disordering according to the Andrich threshold were ‘bridging’ and ‘picking up a pen from the floor.’ (Refer to Appendix H and Table 14) These two items had poor distribution of responses. First, the ‘ability to complete a bridge’ appears to be difficult with discrimination. The areas that are disordered are the categories for ‘moderate/maximum assistance’ and ‘minimum assistance/contact guard.’ There are many factors that can affect bridging in the ICU. One is insertion of lines, the second is patient tolerance and strength, and the third to consider is the bed. Having a good foothold in a bed is a key component to completing a bridge. It is a possibility that patients who were stronger and could maneuver better in the bed could complete the task as opposed to those with less strength and less lower extremity range of motion (i.e., ability to come to hook-lying position).

The second task noted to have disordering was picking up a pen from the floor. Again, the disordered category was between the ‘moderate/maximum assistance’ and ‘minimum assistance/contact guard.’ The difference could lie solely in that the ‘moderate/maximum’ category leaned closer towards the ‘dependent’ assist in scoring due to safety. Concern for a
patient fall may offset a clinical willingness to let the patient complete the task with as little help as possible. Erring on the side of safety may have caused the therapist/data-collector to score the patients lower than actual ability. Both categories should be reassessed within the ICU for more complete evaluation of their usefulness in this setting. As stated previously, clinically, these items have an important role in strength and balance abilities, and so were retained in the assessment tool.

The CPFM-15 has demonstrated its ability to predict discharge from the hospital to home setting compared to other possible settings, on admission and on discharge from the ICU. Fifty-nine percent of the current study sample were discharged home. Utilizing the nomogram in Figure 11 below, the likelihood ratios illustrate how to utilize the results of the findings. If a patient exceeds the admission cut off score (> 42 raw or 51 equal-interval), the nomogram and the LR+ of 2.22 estimates that this individual has a 76% probability of being able to discharge to home. Conversely, a patient who fails to meet or exceed the admission cut-off score has only a 38% probability of being able to discharge to home with an LR- of 0.42.

For a patient who scored at or above the cut-off values on discharge from the ICU, the probability of accurately indicating discharge to home increases. Utilizing the nomogram in Figure 12 below, the likelihood ratios further confirm the certainty that the CPFM-15 is able to predict discharge to the home setting based on discharge scores. If a patient exceeds the discharge cut-off score (>54 raw or 61 equal-interval), the nomogram and the LR+ of 4.53 estimates that this individual has an 87% probability of being able to discharge to home. This increases the probability in accurately predicting discharge to home by almost 30%. Conversely, a patient who fails to meet or exceed the discharge cut-off score has only a 24% probability of being able to discharge to home with an LR – 0.22. While there still were patients discharged to
home that fell below the discharge cut-off score, the probability in accurately predicting discharge to home did show an increase when using the discharge scores rather than admission scores.

While capturing a patient’s ICU admission score can help support the patients’ plan of care and ultimate discharge planning, a patient’s condition and functional status can change significantly within a short period of time. A recent multi-center study conducted by Stelfox et al highlights a new model of care where patients recovering from critical illness can discharge directly to home from the ICU. Stelfox et al revealed discharge directly to home from the ICU did not increase hospital readmission, emergency room visits, or deaths when compared with similar patients who were transitioned to another hospital unit prior to discharge. Having both the ICU admission and ICU discharge scores readily available can promote efficient discharge planning by the therapist. The CPFM-15 supports quick identification of discharge disposition and assists in answering patient and family expectations from day one of their admission.
Figure 11: Likelihood Ratio Nomogram on Admission to the ICU

Nomogram graphical representation of the probability that an individual admitted to the ICU will be able to discharge home. The blue line plots the overall probability, estimated at 59% (based off this sample), who were able to discharge to home. The positive likelihood ratio (LR+) used when the individual met or exceeded the admission cut-off score, determined the participant has a 76% probability of being able to discharge to home. The red line plots the overall probability and the negative likelihood ratio (LR-) used when the individual does not exceed the admission cut-off score, determining that the individual has only a 38% probability of being discharged to home. Nomogram adapted and printed from [http://araw.medc.uic.edu/cgi-bin/testcalc.pl](http://araw.medc.uic.edu/cgi-bin/testcalc.pl).
Figure 12: Likelihood Ratio Nomogram on Discharge from the ICU

Nomogram graphical representation of the probability that an individual admitted to the ICU will be able to discharge home. The blue line plots the overall probability, estimated at 59% (based off this sample), who were able to discharge to home. The positive likelihood ratio (LR+) used when the individual met or exceeded the discharge cut-off score, determined the participant has an 87% probability of being able to discharge to home. The red line plots the overall probability and the negative likelihood ratio (LR-) used when the individual does not exceed the discharge cut-off score, determining that the individual has only a 24% probability of being discharged to home. Nomogram adapted and printed from http://araw.med.edu/cgi-bin/testcalc.pl.
In comparison to the other ICU measures currently available, the SPPB, FSS-ICU, IMS, PFIT, ACIF, and the FAB, have been evaluated for predictive validity in the ability to discharge to home. Predictive validity for the SPPB, FSS-ICU and IMS was conducted in a single study by Parry et al. The sample size for the FSS-ICU and IMS was 66, while the sample size for the SPPB was 23. In the Parry et al study, the SPPB did not have significant predictive ability for discharge to home. The FSS-ICU and IMS did not have significant predictive ability for DC to home based on admission scores, but did have predictive ability using discharge scores. Tipping et al did not assess the IMS admission scores for predictive validity, but supported the ability of the IMS to predict discharge to home using discharge scores.

The PFIT studies for predictive utility for discharge to home were conflicting. Nordon-Craft et al reported that neither admission nor discharge scores of the PFIT predicted discharge to home. However, Denehy et al reported that a higher PFIT score on admission did have predictive validity for discharge to home, which was supported by Parry et al. However, Parry et al also found that higher PFIT scores on discharge are also predictive of discharge to home.

The results for predictive validity of the SPPB, FSS-ICU, IMS, and PFIT were analyzed utilizing logistic regression. The benefit to using logistic regression for predictive validity lies in the concept of “maximum likelihood,” the best odds for accurate prediction are presented. Odds greater than 1.00 indicate the patient likely belongs with the target group. A limitation in the design of utilizing logistic regression for all the above studies, is that cut-off scores were not reported. The positive results indicated “higher scores are associated with discharge to home;” however, what those “higher scores” are was not included in the study results. In contrast to the FSS-ICU, IMS, and SPPB, the CPFМ-15 shows predictive validity of admission and discharge
scores with cut-off values, which improves utility in early discharge planning. The results of these studies are available in Table 19.

### Table 19: Predictive Validity Comparison—Logistic Regression

<table>
<thead>
<tr>
<th>Measure (Score)</th>
<th>Study</th>
<th>Time of Assessment</th>
<th>APACHE II Score Mean (SD)</th>
<th>Number of Participants (n)</th>
<th>Variable Adjusted</th>
<th>Odds Ratio (OR)</th>
<th>Significance (p)</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPFM-15 (15-75)</td>
<td>Peterson et al current study</td>
<td>Admission to ICU</td>
<td>11.76 (7.0)</td>
<td>150</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Peterson et al current study</td>
<td>Discharge from ICU</td>
<td>11.76 (7.0)</td>
<td>150</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>PFIT (0-10)</td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Admission to ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age</td>
<td>1.59</td>
<td>p = 0.004</td>
<td>1.16-2.17</td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>DC from ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age ICU LOS</td>
<td>1.56</td>
<td>p = 0.005</td>
<td>1.15-2.08</td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Admission to ICU</td>
<td>18.8 (6.0)</td>
<td>144</td>
<td>NR</td>
<td>1.20</td>
<td>p = 0.01</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Nordon-Craft et al&lt;sup&gt;98&lt;/sup&gt;</td>
<td>Admission and DC</td>
<td>18.2 (5.5)</td>
<td>34-39</td>
<td>NR</td>
<td>NR</td>
<td>p = 0.087</td>
<td>NR</td>
</tr>
<tr>
<td>FSS-ICU (0-35)</td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Admission to ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age</td>
<td>NR</td>
<td>p = 0.642</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>DC from ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age ICU LOS</td>
<td>1.09</td>
<td>p = 0.013</td>
<td>1.02-1.16</td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Admission to ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age</td>
<td>NR</td>
<td>p = 0.143</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>DC from ICU</td>
<td>21 (7.0)</td>
<td>66</td>
<td>Age ICU LOS</td>
<td>1.54</td>
<td>p = 0.011</td>
<td>1.10-2.13</td>
</tr>
<tr>
<td></td>
<td>Tipping et al&lt;sup&gt;106&lt;/sup&gt;</td>
<td>DC from ICU</td>
<td>19.1 (7.6)</td>
<td>133</td>
<td>Age, APACHE II, FCI</td>
<td>1.16</td>
<td>p = 0.03</td>
<td>1.02-1.32</td>
</tr>
<tr>
<td>SPPB (0-12)</td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Admission and DC</td>
<td>21 (7.0)</td>
<td>23</td>
<td>Age ICU LOS</td>
<td>NR</td>
<td>p &gt; 0.05</td>
<td>NR</td>
</tr>
</tbody>
</table>

**Key:** Areas in **bold** indicate significant “p” values, APACHE II: acute physiology and chronic health evaluation II, ICU: intensive care unit, NR: not reported, LOS: length of stay, N/A: not applicable, CPFM: comprehensive physical function measure, PFIT: physical function in intensive care test, FSS-ICU: functional status score for the ICU, IMS: ICU mobility scale, SPPB: short physical performance battery.
In comparison, there are some differences between this current study and those completed for the SPPB, FSS-ICU, IMS and PFIT. The studies by Parry et al, Tipping et al, and Denehy et al were conducted in Australia and New Zealand. This may limit the generalizability to a U.S. population. Similar to this current study, the study by Nordon-Craft et al was conducted in the United States. The Apache II scores for all of these studies reveal a higher average than the current study, which is likely due to their requirement for mechanical ventilation and the prolonged length of stay (LOS). Denehy and colleagues did not require mechanical ventilation for inclusion, but did purposely target a prolonged ICU LOS by including patients with an expected LOS greater than 5 days. Denehy et al reported their average ICU LOS over the prior 18 months was 2.5 days.

Three studies, Parry et al, Tipping et al, and Nordon-Craft et al, all had the inclusion requirement of mechanical ventilation for a minimum of 48 hours (Nordon-Craft and colleagues required 4 days or longer). Nordon-Craft et al and Tipping et al reported the ICU LOS (median (IQR)) as 11 (6-17) and 20 (12-26) days respectively. This current study was closer in comparison of ICU LOS to Denehy et al and Parry et al with medians (IQR) of 5 (2-55), 7 (6-11), and 8 (5-15), respectively. In contrast to this current study, all four above mentioned studies excluded patients with neurological or neurosurgical diagnoses. The overall increased medical acuity of the patients from the above studies could explain their limited predictive ability for discharge to home utilizing admission scores. The current study for the CPFM did not require patients to be on mechanical ventilation, did include patients with neurological diagnoses, and had a minimum predicted 48-hour ICU LOS requirement. This may indicate the CPFM-15’s usefulness over the FSS-ICU, IMS, SPPB, or PFIT due to its predictive validity with both admission and discharge scores.
Similar to the CPFM-15, predictive validity for both the ACIF\textsuperscript{112} and FAB\textsuperscript{96} were evaluated utilizing ROC analysis for predicting discharge to home. The ACIF was evaluated for predictive validity utilizing discharge scores in an Australian ICU, including a sample of 42 patients, average age (SD) of 59 (19).\textsuperscript{112} The Apache II score on admission was 19 (7). Participants were reviewed for inclusion to the study after day 3 of admission. Bissett et al\textsuperscript{112} reported good predictability of discharge to a place ‘other than home’ with a discharge cut-off score < 0.40 (0-1); AUC 0.79 (0.64-0.89), sensitivity 78%, and specificity 47%.\textsuperscript{112} The predictive values and likelihood ratios were not reported.

The FAB was evaluated for predictive validity in a United Kingdom burn specialty ICU, with a sample size of 56 patients, average age 38.6.\textsuperscript{96} Within that sample, 48 patients were discharged home. The FAB did not report an Apache II score as it is a burn facility. The ICU LOS was (median (IQR)) 12 (7-20) days. The FAB demonstrated good predictability to a place other than home using an admission cut-off score ≤ 9 (7-35) with AUC 0.74 (0.60-0.84), sensitivity 56%, specificity 100%, PPV 100%, and NPV 86%; and a discharge cut-off score ≤ 26 (7-35) with AUC 0.96 (0.87-0.99), sensitivity 75%, specificity 100%, PPV 100%, and NPV 92%.\textsuperscript{96}

Both the ACIF and FAB looked at discharge prediction to a destination ‘other’ than home.\textsuperscript{96,112} In comparison, the current study looked at the predictive ability to discharge to home. Patients who fell below the cut-off value for the FAB were either discharged to a rehab center or sent home with “social care” making differentiation difficult. This current study for the CPFM-15 and the study for the ACIF\textsuperscript{112} both included home care, outpatient PT, or no home services under their differentiation of “home.” The CPFM-15 predictive values are more balanced, with admission score yielding a sensitivity of 71.7%, specificity of 67.7%, PPV of 75.9%, and NPV
of 62.9%, and discharge score yielding a sensitivity of 81.6%, specificity of 82.0%, PPV of 86.8%, and NPV of 75.8%. The ACIF has a low specificity of 47% with discharge scores. Although the FAB showed low sensitivity of 56% with admission scores, the discharge scores did improve. Comparative characteristics between these measures can be seen in Table 20.

Table 20: Predictive Validity Comparison - ROC Analysis

<table>
<thead>
<tr>
<th>Measure (Score)</th>
<th>Suggested Cut off Score (raw points)</th>
<th>Area under the curve (AUC)</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPFM-15 (15-75)</td>
<td>Admit: ≥42</td>
<td>0.773</td>
<td>&lt;0.001</td>
<td>0.700-0.847</td>
<td>71.7%</td>
<td>67.7%</td>
<td>75.9%</td>
<td>62.9%</td>
</tr>
<tr>
<td></td>
<td>DC: ≥54</td>
<td>0.919</td>
<td>&lt;0.001</td>
<td>0.878-0.960</td>
<td>81.6%</td>
<td>82.0%</td>
<td>86.8%</td>
<td>75.8%</td>
</tr>
<tr>
<td>ACIF (0-1)</td>
<td>DC: ≤40</td>
<td>0.79</td>
<td>NR</td>
<td>0.64-0.89</td>
<td>78%</td>
<td>47%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>FAB (7-35)</td>
<td>Admit: ≤9</td>
<td>0.74</td>
<td>NR</td>
<td>0.60-0.84</td>
<td>56%</td>
<td>100%</td>
<td>100%</td>
<td>86%</td>
</tr>
<tr>
<td></td>
<td>DC: ≤26</td>
<td>0.96</td>
<td>NR</td>
<td>0.87-0.99</td>
<td>75%</td>
<td>100%</td>
<td>100%</td>
<td>92%</td>
</tr>
</tbody>
</table>

Key: Admit = Admission, DC = Discharge, Sig = significance, PPV = positive predictive value, NPV = negative predictive value, NR = not reported, CPFM = comprehensive physical function measure, ACIF = acute care index of function, FAB = functional assessment of burns

Patient characteristics in this study are comparable to other studies conducted in the ICU. Wunsch et al looked at three-year outcomes for Medicare beneficiaries who survive their intensive care unit stay. That study demonstrated the overall hospital length of stay for medical and surgical ICU patients was five to seven days. Wunsch et al reported 53.2% of the ICU survivors were discharged home, with 33% being discharged to an inpatient rehabilitation facility. The Wunsch et al study revealed a higher post-ICU discharge mortality for those transferred to a skilled care facility versus those who were discharged to home. Those discharged to a skilled care facility had a 6-month mortality of 24.1% as compared with those discharged to home (7.5%). In the first year after discharge from the hospital that included an ICU stay, re-hospitalizations were 36.1%.
Table 21: Comparison of Patient Characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Participants (n)</th>
<th>ICU LOS (days)</th>
<th>Hospital LOS (days)</th>
<th>Time from Order to PT evaluation (days)</th>
<th>APACHE II Mean (SD)</th>
<th>DC Disposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Dissertation Study</td>
<td>157</td>
<td>Mean 7.14 Median 5.0</td>
<td>Mean 9.8 Median 7.0</td>
<td>1.15</td>
<td>11.76 (0-36)</td>
<td>Home 58.6% Rehabilitation 35.1%</td>
</tr>
<tr>
<td>Wunsch et al&lt;sup&gt;8&lt;/sup&gt;</td>
<td>35 308</td>
<td>Median 5-7</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Lai et al&lt;sup&gt;90&lt;/sup&gt;</td>
<td>90</td>
<td>Mean 8.2</td>
<td>Mean 9.9</td>
<td>&lt; 72hrs</td>
<td>15.6 (±6.1)</td>
<td>NR</td>
</tr>
<tr>
<td>Morris et al&lt;sup&gt;77&lt;/sup&gt;</td>
<td>165</td>
<td>Mean 7.6</td>
<td>Mean 14.9</td>
<td>Days to first OOB 8.5 vs 13.7</td>
<td>21.6 (±8)</td>
<td>Home 72.6% Rehabilitation 20%</td>
</tr>
<tr>
<td>Schweickert et al&lt;sup&gt;74&lt;/sup&gt;</td>
<td>49</td>
<td>Mean 5.9</td>
<td>Mean 13.5</td>
<td>Mean 1.5</td>
<td>20.0 (15.8-24)</td>
<td>Home 43% Rehabilitation 27%</td>
</tr>
<tr>
<td>Routsi et al&lt;sup&gt;78&lt;/sup&gt;</td>
<td>68</td>
<td>Mean 14</td>
<td>NR</td>
<td>NR</td>
<td>18 (±4)</td>
<td>NR</td>
</tr>
<tr>
<td>Denethy et al&lt;sup&gt;79&lt;/sup&gt;</td>
<td>74</td>
<td>Median 7.0</td>
<td>Median 20.0</td>
<td>NR</td>
<td>20.7 (±7.7)</td>
<td>Home 40% Rehabilitation 19%</td>
</tr>
<tr>
<td>Moss et al&lt;sup&gt;80&lt;/sup&gt;</td>
<td>59</td>
<td>Mean 15</td>
<td>Mean 21</td>
<td>Median 8</td>
<td>17.9 (±6.2)</td>
<td>Home 51% Rehabilitation not reported</td>
</tr>
<tr>
<td>Burtin et al&lt;sup&gt;75&lt;/sup&gt;</td>
<td>45</td>
<td>Mean 25</td>
<td>Mean 36</td>
<td>Mean 14</td>
<td>25 (±4)</td>
<td>Home 66% Rehabilitation 17%</td>
</tr>
<tr>
<td>Ragavan et al&lt;sup&gt;109&lt;/sup&gt;</td>
<td>26</td>
<td>NR</td>
<td>Mean 6.8</td>
<td>NR</td>
<td>NR</td>
<td>Home 38.5% Rehabilitation 42.3%</td>
</tr>
<tr>
<td>Hiser et al&lt;sup&gt;164&lt;/sup&gt;</td>
<td>76</td>
<td>Median 6</td>
<td>Median 8</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Nawa et al&lt;sup&gt;100&lt;/sup&gt;</td>
<td>20</td>
<td>Median 4</td>
<td>Median 18</td>
<td>16.5 (7-30)</td>
<td>Home 45% Rehabilitation 15%</td>
<td></td>
</tr>
<tr>
<td>Perme et al&lt;sup&gt;55&lt;/sup&gt;</td>
<td>35</td>
<td>Median 6</td>
<td>Median 14</td>
<td>NR</td>
<td>20 (7-31)</td>
<td>Home 51.4% Rehabilitation 11.4%</td>
</tr>
<tr>
<td>Smailes et al&lt;sup&gt;96&lt;/sup&gt;</td>
<td>56</td>
<td>Median 12</td>
<td>Median 30</td>
<td>NR</td>
<td>NR</td>
<td>Home 86%</td>
</tr>
<tr>
<td>Corner et al&lt;sup&gt;92&lt;/sup&gt;</td>
<td>499</td>
<td>Mean 6.8</td>
<td>Mean 31.8</td>
<td>NR</td>
<td>15 (10-20)</td>
<td>Home 34.3%</td>
</tr>
<tr>
<td>Bissett et al&lt;sup&gt;112&lt;/sup&gt;</td>
<td>42</td>
<td>Mean 19</td>
<td>Mean 51</td>
<td>NR</td>
<td>19 (±7)</td>
<td>Home 35% Rehabilitation 43%</td>
</tr>
<tr>
<td>Tipping et al&lt;sup&gt;106&lt;/sup&gt;</td>
<td>192</td>
<td>Median 11</td>
<td>Median 24</td>
<td>NR</td>
<td>19.1 (7.6)</td>
<td>Home 39.6%</td>
</tr>
<tr>
<td>Study</td>
<td>Number of Participants (n)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ICU LOS (days)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hospital LOS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Time from Order to PT evaluation (days)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>APACHE II Mean (SD)</td>
<td>DC Disposition</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------</td>
<td>----------------------------</td>
<td>--------------------------</td>
<td>---------------------------------------------------</td>
<td>---------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Kasotakis et al&lt;sup&gt;102&lt;/sup&gt;</td>
<td>113</td>
<td>Mean 5.21</td>
<td>Mean 14.1</td>
<td>NR</td>
<td>15.7 (±6.9)</td>
<td>NR</td>
</tr>
<tr>
<td>Nordon-Craft et al&lt;sup&gt;98&lt;/sup&gt;</td>
<td>51</td>
<td>Median 20</td>
<td>Median 26</td>
<td>Mean 15</td>
<td>18.2 (±5.5)</td>
<td>Home 51% Rehab 10%</td>
</tr>
<tr>
<td>Denehy et al&lt;sup&gt;54&lt;/sup&gt;</td>
<td>116</td>
<td>Median 7</td>
<td>Median 23</td>
<td>Median 6</td>
<td>18.8 (±6)</td>
<td>Home 63.8%</td>
</tr>
<tr>
<td>Sommers et al&lt;sup&gt;51&lt;/sup&gt;</td>
<td>115</td>
<td>NR</td>
<td>NR</td>
<td>Mean 6</td>
<td>15.2 (±4.8)</td>
<td>NR</td>
</tr>
<tr>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>66/23&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Median 8</td>
<td>NR</td>
<td>NR</td>
<td>21 (±7)</td>
<td>Home 56% Rehab 30%</td>
</tr>
</tbody>
</table>

Key: <sup>a</sup> = values are reported for study groups, not control groups. <sup>b</sup> = participant number for the short physical performance battery only, ICU = intensive care unit, LOS = Length of stay, PT = physical therapy, APACHE II: Acute Physiology and Chronic Health Evaluation II, (SD): standard deviation, DC = discharge, NR = not reported, N/A = not applicable

Discharge disposition from the hospital is a key concern, but it is one that is multifaceted. Literature supports that while approximately 5% of patients admitted to the hospital are admitted to ICU, they still constitute the largest portion of hospital costs due to expensive equipment, higher number of ICU staff per patient, additional medications, and use of testing. The burden placed on caregivers becomes an issue from both an emotional and cost perspective. It is often a challenge for those bearing this burden to take a family leave or remove themselves from work to care for a loved one.

For example, an independent patient may need discharge to a skilled nursing facility if they are receiving intravenous medication not covered by insurance for administration at home. There are patients requiring assistance that may discharge to home with support of family or self-paid 24-hour care. Early detection of functional concerns or limitations that may prevent a patient from being discharged home is paramount in addressing these social concerns and issues.
during the discharge planning process. While the cut-off score of a measure for discharge to home can assist physical therapy clinicians in addressing patient and family concerns and making relevant, objective discharge recommendation, the CPFM-15 score needs to be considered in context of a patient’s specific social and medical need.

**Relevance to Clinical Practice**

Improving the survivorship experience of patients in the ICU is a defining challenge. Physical function is predictive of length of stay, post-discharge survival, healthcare utilization, quality of life, and return to home. The CPFM-15 developed in this study, highlights areas of mobility according to the WHO’s International Classification of Functioning, Disability, and Health (ICF) model, addressing body function/structural impairments and activity limitations that impact quality of life. The items included in the CPFM-15 encompass mental alertness, neuromuscular/strength assessment, and all functional activities that would need to be completed for basic activities of daily living.

The rehabilitation field has struggled as a whole for a comprehensive and yet clinically sensitive outcome measure for the ICU setting. An added challenge has been in designing a tool that has strong psychometric properties and is efficient and effective to use in an ICU setting. A recent article by Parry et al discussed four major considerations when selecting an instrument. These considerations were the purpose of the assessment, patient capacity, measurement properties, and clinical utility.

Swinkels et al and Jette et al also highlighted the importance of the ‘purpose of the assessment’ and ‘patient capacity’ in their discussion of limitations on why therapists do not utilize outcome measures. The discussion within all three articles highlighted that many
therapists were uncertain as to which tool to apply to the different population of patients. Having measures designed specifically for one patient setting or one diagnosis limits generalizability and causes confusion. In the instance of the previous 14 measures designed specifically for the ICU, there are limitations in the populations for which they have been tested. For example, the FCS\textsuperscript{125,126,166,167} and EFA\textsuperscript{168} were specifically created for a specific neurosurgical population with brain tumors and aneurysms. The FAB\textsuperscript{96} was created for patients in the ICU with burns. The PFIT\textsuperscript{54,97,98} was created for the ICU, but only tested on mechanically ventilated patients. Additionally, all current studies excluded patients with neurological impairment, including stroke or spinal cord injury, and excluded patients with a heart attack or pulmonary embolism within 3 weeks. The Perme\textsuperscript{55,100} has been evaluated solely on cardiovascular patients.

In this current study the CPFM-15, has been shown to assess function across a sample of patients with broad acuity levels and multiple diagnoses to support generalizability across different health care ICU settings, clearly defining the purpose of the assessment, and patient capacity. Testing was conducted across a broad range of neurological, cardiovascular, medical, and surgical groups, including but not limited to, cerebrovascular accidents, open heart surgeries, organ transplants, respiratory failure, and sepsis. The CPFM-15 did not, however, cover areas specific to trauma or burns as this is not an area of expertise of the facility at which this study was conducted.

In terms of measurement properties, the psychometric testing of the prior 14 ICU measures has been detailed previously and can be reviewed in Appendices A-C. An additional concern for three of the 14 measures -ACIF, DEMMI, and SPPB- is that they were developed initially for use in a non-ICU setting.\textsuperscript{36,158} While the DEMMI did undergo rigorous testing,\textsuperscript{114-116} a large difference between creation of the DEMMI and the CPFM-15 is the sample population
and the pool of items. de Morton and colleagues\textsuperscript{114,115} conducted the final Rasch analysis utilizing 106 participants on a general medical ward with patients aged 65 or greater. The mean age in years of the sample for the initial development study was 79.2. The mean age for this study was 65.9. There were nine of the 15 items from the DEMMI retained in the CPFM-15. These included: rolling to the side, bridging, supine to sit, static seated balance, standing on toes, picking up a pen from the floor, sit to stand, assistance needed during ambulation, and ability to jump. In contrast to the DEMMI, the CPFM-15 specifically delineated rolling left and rolling right to specifically evaluate the neurological population.

The strength of the CPFM-15 versus the DEMMI, however, lies in the pool of items extrapolated for the CPFM coming from ICU specific measures. Although the DEMMI was not created with specific ICU tasks in mind, the testing and omission of items that were not appropriate to an acute care setting served as a framework for this current study to identify specific ICU related items inclusion. With regard to function, the CPFM-15 assesses each component of the DEMMI: bed mobility, transfers, seated and standing balance, gait, and jumping, but adds the aspects of mental status, strength, and stairs to complete the functional picture.

The PFIT to PFIT(s) underwent testing in an Australian ICU. The participants totaled 116 for complete admission and discharge data sets. The strength of the CPFM-15 versus the PFITs, is the analysis of items. The PFIT to PFITs did not address the addition of any further items. Denehy et al\textsuperscript{54} solidified the construct validity of the original measure by removing one item but individually these items were not compared with others to account for their ability to capture a patient’s functional ability. The PFITs study completed their admission assessment between days 5-10.\textsuperscript{54} The CPFM study admission score was a mean of 1.15 days (standard deviation 0.76).
While the PFIT and PFITs both address strength and endurance, the CPF-M-15 includes these constructs, while also including bed mobility, transfers, gait, and stairs, once again, completing the picture for a full functional assessment.

The Perme is arguably one of the most expansive ICU outcome measures available with its inclusion of a pain assessment, presence of lines and tubes, presence of IV drips, and use of mechanical ventilation, added to a functional assessment. However, the Perme’s limitations in scoring and potential for high ceiling effects has also been discussed previously, limiting its practical use in the ICU setting. The Perme has limited psychometric testing with only reliability having been reported.\textsuperscript{55,100} The sample sizes for both studies on the Perme have been small with 20-35 participants, and the generalizability is limited due to its sample of cardiovascular patients.\textsuperscript{55,100}

Another measure that looks at medical status items is the CPAx. This measure addresses respiratory function as well as the ability to cough. Since the CPAx was created in England, these two items probably reflect the continental differences between Europe and Australia versus North America.\textsuperscript{92} In Europe and Australia the roles of a physical therapist and respiratory therapist are considered as one. Contrary to practices in North America, ‘physiotherapists’ in Europe and Australia, have a key role in weaning patients from the ventilator and airway clearance management, with early mobilization not as common of a practice.\textsuperscript{84,92,169}

The medical status components from the Perme and CPAx measures were included in this study for IRT Rasch analysis. Analysis revealed concern for a second dimension and these items, aside from mental status, were misfit to the data and so were omitted from the final CPF-M-15 measure. The challenge of measuring functional ability in the ICU relates to the fluctuation of medical status.
Medical complexity can also be due to variations in practice and setting (i.e., what may be a barrier in one facility, or may be a practice in one facility, is not in another). For example, a study by Malone and colleagues\(^5\) in 2015, indicated that community hospitals reported frequent sedation of patients as a barrier to mobility over academic hospitals. Over-sedation may, by default, score a patient lower on mental status. The community hospital also reported less physical therapy consultations for complex ICU patients with comparison to the academic facilities. The study by Malone and colleagues\(^5\) further reported that therapist confidence in progressive mobility was highly influenced by presence of ventilators, lines, tubes, and medications, possibly making these items more therapist specific instead of a true barrier to mobilization and a functional assessment. Another study by Palmieri and colleagues,\(^^{170}\) conducted in 2012 across their own eight facilities, reported that 65% of the therapists employed felt comfortable with progressive mobilization of an intubated patient. While clinical evaluation of the medical complexity items could benefit a practitioner, further testing should be conducted to determine if these items should be included within a physical functional assessment or perhaps scored separately.

Floor and ceiling effects are an important measurement property to assess the recovery trajectories of patients and the intervention’s efficacy. High floor or ceiling effects limit an instrument’s ability to detect change. While the accepted cut-off is \(< 15\%\) for outcome measures,\(^23\) the lower the floor effects the better. The PFIT,\(^{23,54,98}\) CPAx,\(^{92,99}\) IMS,\(^{23,106}\) and SPPB\(^23\) all have high floor effects on admission to the ICU indicating that the items contained in the assessments are too difficult for patients to complete. The ACIF\(^{112}\) was not assessed on admission, but reports high floor effects on discharge from the ICU. These high floor effects indicate that even on discharge from the ICU, patients are having difficulty with task completion.
Conversely, the PFIT also has high ceiling effects on discharge indicating the need for higher-order tasks to capture higher functioning patients. The CPFM-15 has no floor effect and low ceiling effects. The FSS-ICU and DEMMI are comparable to the CPFM-15 in floor and ceiling effects. Table 22 provides a comparison of known ICU admission and discharge floor and ceiling effects.

Table 22: ICU Outcome Measure Floor and Ceiling Effects

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Floor Effect on Admission</th>
<th>Ceiling Effect on Admission</th>
<th>Floor Effect on Discharge</th>
<th>Ceiling Effect on Discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Dissertation Study-CPFM-15</td>
<td>0%</td>
<td>1.9%</td>
<td>0%</td>
<td>3.3%</td>
</tr>
<tr>
<td>PFIT</td>
<td>9.1%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>1.5%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>1.5%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>10.6%&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>21.5%&lt;sup&gt;54&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;54&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;54&lt;/sup&gt;</td>
<td>22.2%&lt;sup&gt;54&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>32.0%&lt;sup&gt;98&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;98&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;98&lt;/sup&gt;</td>
<td>5%&lt;sup&gt;98&lt;/sup&gt;</td>
</tr>
<tr>
<td>CPAx</td>
<td>66.7%&lt;sup&gt;99&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;99&lt;/sup&gt;</td>
<td>13.3%&lt;sup&gt;99&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;99&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>NR&lt;sup&gt;92&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;92&lt;/sup&gt;</td>
<td>3.2%&lt;sup&gt;92&lt;/sup&gt;</td>
<td>0.8%&lt;sup&gt;92&lt;/sup&gt;</td>
</tr>
<tr>
<td>IMS</td>
<td>16.7%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>4.7%&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>96%&lt;sup&gt;106&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;106&lt;/sup&gt;</td>
<td>14%&lt;sup&gt;106&lt;/sup&gt;</td>
<td>4.0%&lt;sup&gt;106&lt;/sup&gt;</td>
</tr>
<tr>
<td>FSS-ICU</td>
<td>3.0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>3.0%&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
<tr>
<td>ACIF</td>
<td>NR&lt;sup&gt;112&lt;/sup&gt;</td>
<td>NR&lt;sup&gt;112&lt;/sup&gt;</td>
<td>23.8%&lt;sup&gt;112&lt;/sup&gt;</td>
<td>0%</td>
</tr>
<tr>
<td>DEMMI</td>
<td>2.6%&lt;sup&gt;51&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;51&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;51&lt;/sup&gt;</td>
<td>2.6%&lt;sup&gt;51&lt;/sup&gt;</td>
</tr>
<tr>
<td>SPPB</td>
<td>82.6%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>56.5%&lt;sup&gt;23&lt;/sup&gt;</td>
<td>0%&lt;sup&gt;23&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Key: PFIT = Physical function in intensive care test, CPAx = Chelsea critical care physical assessment tool, IMS = ICU mobility scale, FSS-ICU = Functional status score for the ICU, ACIF = Acute care index of function, DEMMI = de Morton Mobility Index, SPPB = Short physical performance battery, NR = not reported

The clinical utility and feasibility of any instrument is important. The level of expertise, training, or time required to administer an outcome measure needs to be efficient. Cognition, respiratory status, pain, hemodynamic stability, and the presence of lines or tubes can dictate patient care and may hinder physical progression of the patient, so tasks chosen need to be quick and easy to complete with few props or materials needed. This reduces set-up time,
space needed within the room, and time needed for the patient to complete the assessment. The CPFM-15 can be completed at bedside with or without the use of assistive devices. The assessment can be completed within the constraints of lines and tubes and does account for progression as those items are removed from the patient. The CPFM-15 can provide efficient and effective identification of functional progression and decline and can better assist the physical therapist in clinical decision-making, geared towards treatment and discharge planning.

In further support of clinical utility, the list of barriers revealed by Swinkels et al\textsuperscript{66} and Jette et al,\textsuperscript{65} also included the ‘perception of easy scoring and use by the therapist.’ The ACIF, for example, covers a wide range of tasks including bed mobility, transfers, gait, and steps, excluding only a strength assessment. While one limitation lies in the responsiveness, only three scoring levels (unable, dependent, or independent) the second lies in the scoring, which is complex and difficult to understand. The Perme also has limited responsiveness and a potential for high ceiling effects since its scoring stops at “minimum assistance.” Patients score the same value for functional levels minimum assistance, contact guard, supervision, and independent. The CPFM-15 is a simple addition Likert scale and scores across the functional levels covering dependent, moderate/maximum assistance, contact guard/minimal assistance, supervision, and modified independent/independent.

As noted previously, multiple measures are needed to address the full scope of impairments and functional deficits typically associated with ICU patients. In review of the 14 currently available ICU outcome measures, there are: a) 4 outcome measures assess strength, b) 6 assess bed mobility, c) 9 assess balance, d) 11 assess transfers, e) 10 assess gait, f) 2 assess stairs, and g) 3 assess mental status (Table 23 below). There is no single functional outcome measure that can now be used across the continuum of ICU care.\textsuperscript{158}
This study has identified the impairments and functional tasks needed to provide a comprehensive physical function assessment measure that captures a full spectrum of physical assessment from dependent to independent for ICU patients. The pool of items generated from these currently used ICU measures enabled the ability to capture a full spectrum of function. There is confidence in the CPFM-15’s predictive ability to discriminate discharge to home with both admission and discharge scores. The CPFM-15 is currently the only ICU physical function measure that has met vigorous psychometric testing standards for creation and provides a comprehensive evaluation of the patient in the ICU. The full spectrum of progress can be
measured through the CPFM-15 including assessment of mental status, strength, bed mobility, transfers, balance, gait, and endurance.

The ICU setting is lacking a clear consensus on the most important outcome measurement instrument. This study has significant clinical implications for physical therapists working in the ICU. The CPFM-15 provides clinicians and researchers with a valid, reliable, practical tool to assess functional outcomes of critically ill patients in the ICU. In addition, the CPFM-15 is a unidimensional tool that can track a patient’s mobility progression from bedbound to fully independent.

**Recommendations for Future Research**

The need for standardized outcome measures has been advocated by the WCPT, the CPA, and the APTA. Identification of an appropriate outcome measure to evaluate physical function, however, requires that the instrument be “fit for purpose,” clinimetrically robust, and clinically applicable. To date, functional and clinimetric gaps have been noted in the current ICU outcome measures that limit the identification of one assessment tool to serve as a gold standard. It is proposed that the CPFM-15 could serve that purpose. The CPFM-15 has undergone rigorous development within a diverse ICU patient population. The measure has also demonstrated its ability to provide accurate, relevant information regarding patient progress within the ICU, as well as serve as a predictor in making recommendations regarding patient discharge disposition.

The CPFM-15 has the potential to be safe, quick, and efficient to administer, however, further testing is needed to evaluate this hypothesis. While there were no untoward effects noted during this study, efficiency and safety were not directly measured. The CPFM-15 includes two
items that did have disordering according to the Andrich threshold. The items ‘bridging’ and ‘picking up a pen from the floor’ should be evaluated for adjustments needed to the Likert scale due to concerns noted in the discussion. The item ‘picking up a pen from the floor’ in combination with the item ‘standing on your toes,’ creates a design that could evaluate higher levels of balance enabling identification of potential for falls and safety risks on discharge to home. Identification of balance and falls risk has been shown to be a healthcare priority, with a recent study showing a 17% occurrence of falls post ICU discharge.171

Further testing should be conducted on the CPFM-15 regarding classical test theory principles, such as establishing inter-rater reliability, minimally important difference (MID), and minimally clinically important difference (MCID) scoring. Although the CPFM-15 was developed and tested in an acute care hospital at the ICU level, the potential applications of this instrument are broad. The breadth of the scale allows for potential use across the hospital continuum, i.e., telemetry and medical-surgical units; and quite possibly in the rehab setting on hospital discharge. Further testing should also be conducted to establish its use across the continuum of care within the hospital and then potentially upon discharge from the hospital or with physical therapy provided in the home setting.

Limitations and Delimitations

Limitations

The primary limitation in this study is that it was conducted in a single ICU. Though that setting did cater to a diverse ICU population, it would increase generalizability to have further testing completed beyond this setting and region. Sample size and data collection were limited to a sample of convenience occurring during the routine work schedule of two therapists assigned
to the ICU of the hospital where the study occurred. The sample size and data collected could have been larger with more therapist-performing assessments and could have limited data not collected, particularly on the date of discharge during the days the two therapists were not scheduled to work. All results are based upon the inclusion criteria of a presumed independent prior level of function with direct admission to the ICU. These results could change with inclusion of patients admitted from other units within the hospital, from other hospitals, or even from other skilled nursing facilities or rehab centers.

**Delimitations**

Sample bias may exist within the data as only patients admitted directly to the ICU were evaluated for inclusion in the study. There were additional patients present from other units within the hospital or as transfers from other facilities that were independent prior to transfer to the ICU; however, their mobility, or lack of mobility, could not be controlled for, and as such, limited the inclusion criteria to those patients with a direct admission to the ICU. A second possible bias was the patient’s ability to sign for consent or to have a family member/POA available to sign consent. This could have limited those patients who were cognitively impaired from participating in the study.

This study was designed for two clinical physical therapists to collect data, but only the primary researcher entered and analyzed the data. While every intention was made to ensure accurate entry of data, there was no secondary data entry check in place. IRT Rasch analysis on a whole, however, does limit bias with its ability to analyze consistency of item responses and patterns in relation to the total score.\(^56\)
Summary

This chapter provides a discussion of the study results, limitations, and recommendations. Currently, multiple measures are needed to address the full scope of impairments and functional deficits typically associated with ICU patients. The lack of such a measure can increase time of administration and may cause repetition in tasks administered. A standardized physical function assessment measure is considered necessary to solidify veracity in research and to promote efficacy in clinical practice. Use of IRT and Rasch analysis lend credibility to this assessment tool as a robust, reliable, and valid comprehensive physical function measure for physical therapists assessing and treating patients in the ICU setting. Future research is recommended to address the use of the CPFM-15 across other settings and within varied patient samples. The ability to communicate across settings with one comprehensive outcome measure could potentially serve as a gold standard for physical therapy assessment.
## Appendix A: ICU Specific Article Characteristics

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Author/Year/Location</th>
<th>Subjects (n)</th>
<th>Age/Gender (mean ± SD)</th>
<th>Setting</th>
<th>Severity Score</th>
<th>ICU LOS (Median/IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Function in Intensive Care Unit (PFIT) and scored (PFITs)</td>
<td>Skinner et al, 2009 Australia</td>
<td>10-12</td>
<td>56.8 (± 12.5) 7 males</td>
<td>Med-Surg, ICU and Resp wean unit</td>
<td>APACHE II 10-25 (6 missing)</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Deney et al, 2013 Australia</td>
<td>116</td>
<td>59.3 (± 15.4) 60% male</td>
<td>ICU</td>
<td>APACHE II 18.8 (6.0)</td>
<td>7 (6-11)</td>
</tr>
<tr>
<td></td>
<td>Nordon-Craft et al, 2014 USA</td>
<td>34-39</td>
<td>50.5 (±16.3) 63% male</td>
<td>ICU</td>
<td>APACHE II 18.2 (5.5)</td>
<td>20 (12-26)</td>
</tr>
<tr>
<td></td>
<td>Parry et al, 2015 Australia</td>
<td>66</td>
<td>58 (±17) M/F not reported</td>
<td>Mixed ICU</td>
<td>APACHE II 21 (±7)</td>
<td>8 (5-15)</td>
</tr>
<tr>
<td>Chelsea Critical Care Physical Assessment Tool (CPAx)</td>
<td>Corner et al, 2013</td>
<td>33</td>
<td>67 (51-75) M/F 25/8</td>
<td>General and Trauma ICU</td>
<td>APACHE II 20 (6)</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Corner et al, 2015 England</td>
<td>30</td>
<td>47.1 (±21.2) 63.3% male</td>
<td>Burn ICU</td>
<td>BOBI score 4 (2.1)</td>
<td>17 (2-94)</td>
</tr>
<tr>
<td></td>
<td>Corner et al, 2014 England</td>
<td>499</td>
<td>62.3 (18.31) M/F not reported</td>
<td>Mixed Med-Surg ICU</td>
<td>APACHE II 16 (10-20)</td>
<td>6 (4-12)</td>
</tr>
<tr>
<td>Perme ICU Mobility Score</td>
<td>Perme et al, 2014 USA</td>
<td>35</td>
<td>67 (26-92) 60% male</td>
<td>CV ICU</td>
<td>APACHE II 20 (7-31)</td>
<td>6 (1-24)</td>
</tr>
<tr>
<td></td>
<td>Nawa et al, 2014 USA</td>
<td>20</td>
<td>64.5 (20-86) 60% male</td>
<td>CV ICU</td>
<td>APACHE II 16.5 (7-30)</td>
<td>4 (1-42)</td>
</tr>
<tr>
<td>Surgical Intensive Care Unit Optimal Mobilization Score (SOMS)</td>
<td>Kasotakis et al, 2012 USA</td>
<td>113</td>
<td>60.2 (18-93) 58.4% male</td>
<td>Surgical ICU</td>
<td>APACHE II 15.68 (3-32)</td>
<td>5.21 (1-31)</td>
</tr>
<tr>
<td></td>
<td>Schaller et al, 2016 Germany</td>
<td>128</td>
<td>Non-neurocritical 63 (24-89) 68% male</td>
<td>Surgical ICU</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Author/Year/Location</td>
<td>Subjects (n)</td>
<td>Age/Gender (mean ± SD)</td>
<td>Setting</td>
<td>Severity Score</td>
<td>ICU LOS (Median/IQR)</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Piva et al, 2015</td>
<td>98</td>
<td>61.5 (SD 16.4) 68.3% male</td>
<td>General and Neuro ICU</td>
<td>SAPS II 41.1 (15.7)</td>
<td>8 (4-13.5)</td>
</tr>
<tr>
<td></td>
<td>Garzon-Serrano et al, 2011</td>
<td>63</td>
<td>RN group 58 (± 15) 66% male  PT group 58 (± 15) 66% male</td>
<td>Surgical ICU</td>
<td>RN group APACHE II 9 ± 8</td>
<td>PT group APACHE II 12 ± 12</td>
</tr>
<tr>
<td>ICU Mobility Scale (IMS)</td>
<td>Tipping et al, 2016</td>
<td>192</td>
<td>58 (±15.8) 61% male</td>
<td>Mixed ICU</td>
<td>APACHE II 19.1 (SD 7.6)</td>
<td>11 (6-17)</td>
</tr>
<tr>
<td></td>
<td>Hodgson et al, 2014</td>
<td>100</td>
<td>58 (±17) 38% male</td>
<td>Mixed Med-Surg and Trauma ICU</td>
<td>APACHE II 19 (±7)</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Parry et al, 2015</td>
<td>66</td>
<td>58 (±17) 61% male</td>
<td>Mixed ICU</td>
<td>APACHE II 21 (±7)</td>
<td>8 (5-15)</td>
</tr>
<tr>
<td>Functional Status Score for the ICU (FSS-ICU)</td>
<td>Ragavan et al, 2016</td>
<td>26</td>
<td>53.65 (24-88) 46.2% male</td>
<td>Mixed ICU</td>
<td>Not reported</td>
<td>4.04 (1-24)</td>
</tr>
<tr>
<td></td>
<td>Parry et al, 2015</td>
<td>66</td>
<td>58 (±17) 61% male</td>
<td>Mixed ICU</td>
<td>APACHE II 21 (±7)</td>
<td>8 (5-15)</td>
</tr>
<tr>
<td>Acute Care Index of Function (ACIF)</td>
<td>Bissett et al, 2016</td>
<td>42</td>
<td>59 (16-88) 74% male</td>
<td>Mixed Med-Surg ICU</td>
<td>APACHE II 19 (±7)</td>
<td>19 (2-119)</td>
</tr>
<tr>
<td>De-Morton Mobility Index (DEMMI)</td>
<td>Sommers et al, 2016</td>
<td>115</td>
<td>61 (±16.1) 67% male</td>
<td>Mixed Med-Surg ICU</td>
<td>APACHE II 15.2 (4.8) SOFA score 7 (3.6)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Short Physical Performance Battery (SPPB)</td>
<td>Parry et al, 2015</td>
<td>(66) But only n = 23 for SPPB</td>
<td>58 (±17) 61% male</td>
<td>Mixed ICU</td>
<td>APACHE II 21 (±7)</td>
<td>8 (5-15)</td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Author/Year/Location</td>
<td>Subjects (n)</td>
<td>Age/Gender (mean ± SD)</td>
<td>Setting</td>
<td>Severity Score</td>
<td>ICU LOS (Median/IQR)</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Early Functional Abilities Scale (EFA)</td>
<td>Alvsåker et al,124 2011 Norway</td>
<td>24</td>
<td>25 (20-37) 79% male</td>
<td>Trauma and Surgical ICU. Specific TBI patients</td>
<td>Not reported</td>
<td>15 (9-19)</td>
</tr>
<tr>
<td>Functional Capacity Scale (FCS)</td>
<td>Ślusarz et al,167 2012 Poland</td>
<td>97</td>
<td>51 (±14) 33.6% male</td>
<td>Surgical ICU. Specific intracranial aneurysm</td>
<td>Hunt and Hess Scale</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Ślusarz et al,125 2014 Poland</td>
<td>23 &amp; 97</td>
<td>Age 21-40 (17.4%) Age 41-60 (52.2%) Age &gt; 60 (30.4%) 39.1% male overall</td>
<td>Surgical ICU. Specific intracranial aneurysm</td>
<td>Hunt and Hess Scale</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Ślusarz,166 2007 Poland</td>
<td>40</td>
<td>50.4 (±16.8) 50% male</td>
<td>Surgical ICU Specific for brain neoplasm surgery</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td></td>
<td>Ślusarz et al,126 2014 Poland</td>
<td>159</td>
<td>Male average age: 55 Female average age 64 71.7% male</td>
<td>ICU specific for TBI</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Medical Research Council Sum Score (MRC-SS) and Hand-Held</td>
<td>Hermans et al,94 2012 Belgium</td>
<td>MRC-SS: 75 HHD: 46</td>
<td>MRC-SS 59 (52-71) 50.7% male HHD *(47-68) 58.7% male</td>
<td>Surgical and Medical ICU</td>
<td>Not reported</td>
<td>MRC-SS 22 (15-30) HHD 15 (9-32)</td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Author/Year/Location</td>
<td>Subjects (n)</td>
<td>Age/Gender (mean ± SD)</td>
<td>Setting</td>
<td>Severity Score</td>
<td>ICU LOS (Median/IQR)</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Dynamometry (HHD)</td>
<td>Hough et al,24 2011 USA</td>
<td>30 (only 10 in ICU)</td>
<td>49 ±15 71% male</td>
<td>ICU</td>
<td>Not reported</td>
<td>Not Reported</td>
</tr>
<tr>
<td>Connolly et al,128 2013 England</td>
<td>For Reliability 20 For Predictive 49-65</td>
<td>For Reliability 67.5 (51.8-75) 12 males For Predictive 66.0 (54.8-76.3) 64 males</td>
<td>Mixed Med-Surg ICU</td>
<td>Reliability APACHE II 19.5 (15.5-24.0) Predictive APACHE II 17.0 (15.0-22.0)</td>
<td>Reliability 33.5 (25.5-58) Predictive 11.0 (6.0-25.3)</td>
<td></td>
</tr>
<tr>
<td>Ali et al,130 2008 USA</td>
<td>For Predictive 136 For Reliability 12</td>
<td>57.7 (± 15.5) 47.8% male</td>
<td>Medical ICU</td>
<td>APACHE III 65.8 (±27) SOFA 6.4 (±3.0)</td>
<td>ICUAP 21 (±11) No ICUAP 12 (±6)</td>
<td></td>
</tr>
<tr>
<td>Baldwin et al,131 2013 Australia</td>
<td>17</td>
<td>78 (46-82) 59% male</td>
<td>ICU</td>
<td>APACHE II 20 (5) SOFA 2(2)</td>
<td>18 (12-21)</td>
<td></td>
</tr>
<tr>
<td>Vanpee et al,129 2011 Belgium</td>
<td>39</td>
<td>64 (53-72) 62% male</td>
<td>Med-Surg ICU</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Lee et al,95 2012 USA</td>
<td>95</td>
<td>61.2 (18.3) 59% male</td>
<td>Surg ICU</td>
<td>APACHE II 15.2 (9.0)</td>
<td>5 (3-9.5)</td>
<td></td>
</tr>
<tr>
<td>Functional Assessment for Burns (FAB)</td>
<td>Smailes et al,96 2013 United Kingdom</td>
<td>56</td>
<td>38.6 (34.3-43) M/F not reported</td>
<td>Burns ICU</td>
<td>TBSA 35% (30-45)</td>
<td>12 (7-19.65)</td>
</tr>
</tbody>
</table>

KEY: *article records a 4 for median age of patient’s but reports IQR or (47-68). Number of subjects discrepancy: Some authors used a different number of subjects for the psychometric testing, this has been noted. Abbreviations: Med = medical, Surg = surgical, APACHE = Acute Physiology and Chronic Health Evaluation, SOFA = Sequential Organ Failure Assessment, TBSA = Total Burn Surface Area, ICUAP = ICU Acquired Weakness, CV ICU = cardiovascular ICU.
### Appendix B: Cosmin Scores

<table>
<thead>
<tr>
<th>Measure</th>
<th>Article</th>
<th>Itemized Response</th>
<th>Internal Consistency</th>
<th>Reliability Box B</th>
<th>Measurement Error Box C</th>
<th>Content Validity Box D</th>
<th>Structural Validity Box E</th>
<th>Hypothesis Testing Box F</th>
<th>Cross-Cultural Validity Box G</th>
<th>Criterion Validity Box H</th>
<th>Responsive Box I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PFIT &amp; PFIT-s</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor - Excellent</td>
<td>Skinner et al&lt;sup&gt;97&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Denehy et al&lt;sup&gt;54&lt;/sup&gt;</td>
<td>E</td>
<td>E</td>
<td></td>
<td>E</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P/F*</td>
</tr>
<tr>
<td></td>
<td>Nordon-Craft et al&lt;sup&gt;98&lt;/sup&gt;</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P/P/P</td>
</tr>
<tr>
<td><em><em>CPA</em> x</em>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor - Excellent</td>
<td>Corner et al&lt;sup&gt;53&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td></td>
<td>E</td>
<td>F</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>Corner et al&lt;sup&gt;99&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>Corner et al&lt;sup&gt;92&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td>G</td>
</tr>
<tr>
<td><strong>Perme Mobility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>Perme et al&lt;sup&gt;55&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td></td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nawa et al&lt;sup&gt;100&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FSS-ICU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor - Fair</td>
<td>Ragavan et al&lt;sup&gt;109&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parry et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td><strong>SOMS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>Kasotakis et al&lt;sup&gt;102&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schaller et al&lt;sup&gt;103&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>P</td>
</tr>
<tr>
<td>Measure</td>
<td>Article</td>
<td>Itemized Response Theory</td>
<td>Itemized</td>
<td>Internal Consistency</td>
<td>Reliability</td>
<td>Measurement Error</td>
<td>Content Validity</td>
<td>Structural Validity</td>
<td>Hypothesis Testing</td>
<td>Cross-Cultural Validity</td>
<td>Criterion Validity</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>--------------------------</td>
<td>----------</td>
<td>----------------------</td>
<td>-------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>IMS</td>
<td>Piva et al&lt;sup&gt;104&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Garzon-Serrano et al&lt;sup&gt;101&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Tipping et al&lt;sup&gt;106&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Hodgson et al&lt;sup&gt;105&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Parry et al&lt;sup&gt;123&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Bissett et al&lt;sup&gt;112&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Sommers et al&lt;sup&gt;151&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Parry et al&lt;sup&gt;123&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Alvsåker et al&lt;sup&gt;124&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Ślusarz et al&lt;sup&gt;167&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Ślusarz et al&lt;sup&gt;125&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>IMS</td>
<td>Ślusarz et al&lt;sup&gt;166&lt;/sup&gt;</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
</tbody>
</table>

159
<table>
<thead>
<tr>
<th>Measure</th>
<th>Article</th>
<th>Itemized Response Theory</th>
<th>Itemized Consistency Box A</th>
<th>Reliability Box B</th>
<th>Measurement Error Box C</th>
<th>Content Validity Box D</th>
<th>Structural Validity Box E</th>
<th>Hypothesis Testing Box F</th>
<th>Cross-Cultural Validity Box G</th>
<th>Criterion Validity Box H</th>
<th>Responsive Box I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Šlusarz et al.¹²⁶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MRC-SS/HHD</strong></td>
<td>Hermans et al.²⁴</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor – Good</td>
<td>Hough et al.²⁴</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Connolly et al.²⁸</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ali et al.¹³⁰</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Baldwin et al.¹³¹</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vanpee et al.¹²⁹</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lee et al.¹²⁹</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FAB</strong></td>
<td>Smailes et al.¹⁹⁶</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kappa: $k = 0.976$, Significance $= 0.000$

Key: P = Poor, F = Fair, G = Good, E = Excellent

* Rater 1 & 2 disagreement, third rater reported a Poor.

Physical Function in Intensive care unit Test (PFIT), Chelsea critical care Physical Assessment tool (CPAx), Perme ICU mobility score (Perme), Surgical intensive care unit Optimal Mobilization Score (SOMS), ICU Mobility Scale (IMS), Functional Status Score for the ICU (FSS-ICU), Acute Care Index of Function (ACIF), De-Morton Mobility Index (DEMMI), Short Physical Performance Battery (SPPB), Early Functional Abilities scale (EFA), Functional Capacity Scale (FCS), Medical Research Council Sum Score (MRC-SS), Hand Held dynamometry (HHD), Functional Assessment for Burns (FAB)
### Appendix C: Physical Function Test Psychometric Properties

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness/Predictive Properties</th>
<th>Floor/Ceiling</th>
<th>MCID</th>
<th>Cross Cultural Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFIT &amp; PFITs</td>
<td>Inter-rater (n =10) ICC 0.97-1.00 SEM 0.61 &amp; 0.76</td>
<td>Convergent Validity: <strong>TUG</strong>: $r = -.60$ <strong>6MWT</strong>: $r = .41$ <strong>MRC-SS</strong>: rho = .49</td>
<td>-Mean difference of 86.3 more steps (95% CI, 15.8-156.8; p 0.02) -Increased marching time by a mean difference of 56 sec. (95% CI, 5.2-102.8; p 0.03) -Increased cadence by a mean of 25.4 steps/min (95% CI, -1.7-50.3; p 0.04) -Increased shoulder flexion reps by a mean difference of 8 (CI 95%, 0.5-25.4; p 0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denehy et al,54 2013</td>
<td></td>
<td></td>
<td>ESI: 0.82, 95% CI 0.66-0.99</td>
<td></td>
<td>1.5 points</td>
<td></td>
</tr>
<tr>
<td>Nordon-Craft et al,98 2014</td>
<td></td>
<td></td>
<td>-Initial DC responsiveness (n=26) 1.14. -5.67 (2.01) days post DC from ICU responsiveness 0.59</td>
<td>-32% floor effect on admission -5% floor effect for DC PFIT -5% ceiling effect on discharge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Parry et al,2014</td>
<td></td>
<td></td>
<td>Predictive: higher awakening PFIT scores predicted DC home OR 1.59 p 0.004 Effect Size: 0.71</td>
<td>PFIT awakening floor 9.1% Ceiling 1.5%</td>
<td>1.4 points</td>
<td></td>
</tr>
<tr>
<td>CPAX</td>
<td>Corner et al,2013</td>
<td>Inter-rater reliability K = 0.988, 95% CI 0.791 – 1.000, p &lt; 0.01 α = 0.798 (n = 15)</td>
<td>-Face and content validity Focus group → questionnaire → pilot test CVI 1.00, p &lt; 0.05 -Construct Validity Peak cough flow 0.633, p 0.006, AusTOMs 0.693–0.903, p &lt; 0.001, MRC-SS 0.650, p &lt;0.001, SF-36 (PC) 0.843, p 0.009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corner et al,99 2015</td>
<td></td>
<td></td>
<td></td>
<td>66.7% floor on admission 13.3% floor on discharge No patients scored a full 50</td>
<td>MCID 6 MCD 3</td>
<td></td>
</tr>
<tr>
<td>Corner et al,92 2014</td>
<td></td>
<td></td>
<td>Construct validity: Discharge groups p &lt; 0.0001</td>
<td></td>
<td>0.8% ceiling 3.2% floor</td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Perme</td>
<td>Perme et al, 55 2014</td>
<td>% agreement 94.29% (68.57%-100%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nawa et al, 100 2014</td>
<td>Kappa 0.98 (0.60 – 1.0)</td>
<td>% agreement 80-100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOMS</td>
<td>Kasotakis et al, 102 2012</td>
<td>Kappa 0.80, 95% CI 0.702-0.898</td>
<td>SOMS was predictive of ICU and hospital LOS (p &lt; .001)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SOMS was predictive of ICU and hospital LOS (p < .001)
<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Reliability</th>
<th>Validity</th>
<th>Responsiveness/Predictive Properties</th>
<th>Floor/Ceiling</th>
<th>MCID</th>
<th>Cross Cultural Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schaller et al, 2016</td>
<td>Nurse to Nurse Kappa 0.41/65.6% Nurse to PT Kappa 0.26-0.37/51.4%-57.1%</td>
<td></td>
<td>SOMS was predictive of ICU LOS &amp; hospital LOS (p &lt; .001), and mortality (p .001)</td>
<td></td>
<td></td>
<td>Translated to German</td>
</tr>
<tr>
<td></td>
<td>Nurse to intensivist 0.31-0.35/58%-58.0% PT to intensivist 0.39/57.7% Nurse to achieved SOMS 0.49-0.55/67.6%-70.3% PT to achieved SOMS 0.48/62.4% Intensivist to achieved SOMS 0.53/68.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Piva et al.(^{104}) 2015</td>
<td>SOMS improvement associated with lower hospital mortality (odds ratio 0.06) but an increase in hospital LOS (odds ratio 1.21) The first morning SOMS (ICU admit) indicated better mobility, and was associated with lower ICU and hospital LOS (ratios .89 and .84)</td>
<td></td>
<td></td>
<td></td>
<td>Translated to Italian</td>
<td></td>
</tr>
<tr>
<td>Garzon-Serrano et al.(^{101}) 2011</td>
<td>Face Validity: Internal consensus panel Patient’s level of mobilization achieved was significantly higher than RN’s (p &lt; .001) 73% of RN’s limited treatment to in bed (compared to 37% of PT) 38% of PT’s reached of level of standing and ambulation (13% of RN’s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMS Tipping et al.(^{106}) 2016</td>
<td>Convergent Validity: <strong>MRC-SS</strong> (r = 0.64, 95% CI 0.49-0.75 p &lt;0.001 (n=87)) Predictive: “Alive at 90 days” OR 1.38, 95% CI 1.14-1.66 p &lt;0.001 (n = 154) “discharged home” OR 1.16, 95% CI 1.02-1.32 p 0.03 (n = 133) “return to work” OR 1.09, 95% CI 0.92-1.28 p 0.33 (N= 73) Responsiveness: ESI</td>
<td>96% floor on admission IMS 14% floor at ICU discharge 4% ceiling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Hodgson et al, 2014</td>
<td>Overall ICC 0.80 (0.75-0.84)</td>
<td>Face Validity: Multi-disciplinary group then external expert input via a conference, then international circulation and web posting to researchers specializing in ICU rehabilitation Further content validity: feasibility survey (n=30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parry et al, 2014</td>
<td>Criterion validity + correlation on awakening: to PFITs rho = 0.81 95% CI 0.70-0.88, p &lt; 0.005 on ICU discharge to PFITs rho 0.66 95% CI 0.49-0.80, p &lt; 0.005 (n=64)</td>
<td>Determination of DC home not significant for awakening IMS p 0.143 but significant correlation with higher IMS scores on DC OR 1.54, p 0.011 Effect size: 0.59</td>
<td></td>
<td>IMS awakening score 16.7% floor 0% ceiling IMS Discharge score 0% floor 4.7% ceiling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSS-ICU</td>
<td>ICC 0.992 95% 0.984-0.996. p 0.00 α 0.992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Parry et al, 2014</td>
<td>Convergent validity to PFITs Large + relationship on awakening rho 0.87 95% CI 0.79-0.92, p &lt; 0.005 &amp; ICU discharge rho 0.85 95% CI 0.77-0.90 p &lt; 0.005</td>
<td>Admission FSS-ICU not significant for determining home p 0.642 but significant ICU DC FSS-ICU scores for determining home OR 1.09 p 0.013</td>
<td>On awakening 3% floor 0% ceiling</td>
<td>4.3-5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACIF</td>
<td>Inter-rater</td>
<td>Strong correlation with IMS score r 0.84 p 0.01</td>
<td>ACIF score &lt; 0.40 predicted DC to destination other than home. ROC 0.79 95% CI 0.64-0.89 Sensitivity: 0.78 Specificity: 0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sommers et al, 2016</td>
<td>Inter-rater ICC 0.93 (0.91-0.95)</td>
<td>Convergent validity admission With BI: rho 0.56 With Katz ADL: 0.45 With MRC-SS: 0.57</td>
<td>admission Floor 2.6% Ceiling 0% Discharge Floor 0% Ceiling 2.6%</td>
<td>MDC$_{90}$ 6.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>SPPB</td>
<td>Parry et al, 2014</td>
<td>Convergent with PFIT Moderate + relationship on awakening rho 0.70 95% CI 0.47-0.83, p &lt; 0.005 on ICU discharge large + relationship rho 0.86 95% CI 0.73-0.91 p &lt; 0.005</td>
<td>No difference noted on awakening score and DC scores with regards to DC home p &gt; 0.05 Effect size 0.33</td>
<td>On awakening 82.6% floor 0% ceiling On ICU DC 56.5% floor 0% ceiling</td>
<td>1.5-1.3</td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>--------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>EFA</td>
<td>Alvsåker et al, 2011</td>
<td>Overall agreement between PT and OT K= 0.76 (0.48-0.89) Overall agreement between nursing and MD K= 0.60 (0.22-0.80)</td>
<td></td>
<td>Continence: 70% floor Movement changes and transfers: 42% floor Wakefulness: 42% floor Standing: 35% floor Comprehension: 30% floor Tonus: 30% ceiling Arbitrary movements: 38% ceiling Head control: 30% ceiling Tactile information: 30% ceiling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>FCS</td>
<td>Ślusarz et al, 2012</td>
<td>Spearman Correlations FCS with GOS $r = 0.89$ FCS with FIR $r = 0.93$ FCS with BI $r = 0.82$ FCS with RS $r = -0.88$ Kruskal Wallace FCS with Hunt-Hess score 35.78; $p &lt; 0.001$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ślusarz et al, 2014</td>
<td>W-Kendall Pre-op $0.706-1.000$ Variability per category ($p = 0.0005$)</td>
<td>Spearman Correlation FCS &amp; GOS ($r = 0.890$) $p &lt; .01$ FCS &amp; FIR ($r = 0.932$) $p &lt; .001$ FCS &amp; BI ($r = 0.821$) $p &lt; .001$ FCS &amp; RS ($r = -0.881$) $p &lt; .01$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>---------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Ślusarz et al, 166 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spearmans Correlation FCS and FIR r = 0.78 p &lt; 0.001 FCS and GOS r = -0.049 p &lt; 0.001 FCS and KPS r = 0.56 p &lt; 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ślusarz et al, 126 2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FCS with GCS on admission R = 0.652 p &lt; 0.001 On discharge R = -0.687 p &lt; 0.001 FCS with GOS R = -0.784 p = 0.000</td>
<td>Age groups with low significance with FCS on admission R = 0.261 p &lt; 0.001 On discharge R = 0.140, p = 0.088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>MRC-SS &amp; HHD</td>
<td>Hermans et al, 2012</td>
<td>Reproducibility MRC-SS all limbs ICC 0.95 (0.92-0.97) Weighted kappa 0.83 ±0.03 Hand grip strength R: ICC 0.93 (0.86-0.97) L: ICC 0.97 (0.94-0.98) See article for individual muscle agreement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hough et al, 2011</td>
<td>MRC-SS total inter-rater ICC 0.83 (0.67-0.93) See article for individual muscle group agreement which was poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Connolly et al, 2013</td>
<td>Inter-rater ICC 0.94 (0.85-0.98) Agreement for Dx of ICUAW K= 0.60 (0.25-0.95) See article for individual muscle group agreement</td>
<td>No association between MRC-SS and hospital and ICU mortality (p 0.53 &amp; 0.67) + association with MRC-SS and ICU and Hospital LOS (p 0.004 &amp; 0.04) ICU LOS Sensitivity 92.9% Specificity 40.5% PPV: 54.2% NPV: 88.2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ali et al, 2008</td>
<td>Agreement for identification of ICUAW ICC 0.96, p 0.001</td>
<td>-Hand grip strength compared to MRC-SS for ICUAW identification Sensitivity 80.6% Specificity 83.2% PPV: 63% NPV: 92.3% -Hospital mortality higher with ICUAW OR 7.8 (95% CI 2.4-25.3) p = 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Baldwin et al, 2013</td>
<td>Inter-rater reliability Hand grip L: ICC 0.892 R: ICC 0.924 Elbow flexion L: ICC 0.616 R: ICC 0.714 Knee Extension L: ICC 0.788 R: ICC 0.841 Test-retest Hand grip L: ICC 0.855 R: ICC 0.918 Elbow flexion L: ICC 0.423 R: ICC 0.819 Knee Extension L: ICC 0.909 R: ICC 0.896</td>
<td></td>
<td></td>
<td></td>
<td>ICU patients MDD95 Grip: 20.8% Elbow flexion: 18.5% Knee extension: 19.5%</td>
<td></td>
</tr>
<tr>
<td>Vanpee et al, 2011</td>
<td>Inter-rater agreement ICC for individual muscle groups ranged 0.76-0.96 P = &lt; .0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Measure</td>
<td>Reliability</td>
<td>Validity</td>
<td>Responsiveness/Predictive Properties</td>
<td>Floor/Ceiling</td>
<td>MCID</td>
<td>Cross Cultural Validity</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>------</td>
<td>-------------------------</td>
</tr>
</tbody>
</table>
| Lee et al. 2012 |             |          | - MMT and mortality OR 0.94 (95% CI 0.89-0.98) p .006  
                     - MMT and ICU LOS r = -0.031, (95% CI -0.48- -0.12) p 0.002  
                     - MMT and Hospital LOS r = -0.33 (-0.50 - -0.13) p .001  
                     - Grip strength and mortality OR 1.0 (95% CI 0.93-1.04) p .03  
                     - Grip strength and ICU LOS r = -0.06 (95% CI -0.26- -0.14) p 0.55  
                     - Grip strength and hospital LOS r = -0.05 (95% CI -0.25-0.16) p = 0.65 | | | | |
**Predictive Value**

To home FAB score ≤ 9 on ICU discharge will most likely need support services.
- 56% sensitivity
- 100% specificity
- AUC 0.74 (0.6-0.84)
  +PV 100%
  -PV 86%

Predictive value to home FAB scores ≤ 26 on hospital discharge will DC home.
- 75% sensitivity
- 100% specificity
- AUC 0.96 (0.87-0.99)
  +PV 100%
  -PV 92%

Abbreviations: Physical Function in Intensive care unit Test (PFIT), Chelsea critical care Physical Assessment tool (CPAx), Perme ICU mobility score (Perme), Surgical intensive care unit Optimal Mobilization Score (SOMS), ICU Mobility Scale (IMS), Functional Status Score for the ICU (FSS-ICU), Acute Care Index of Function (ACIF), De-Morton Mobility Index (DEMMI), Short Physical Performance Battery (SPPB), Early Functional Abilities scale (EFA), Functional Capacity Scale (FCS), Medical Research Council Sum Score, (MRC-SS), Hand Held dynamometry (HHD), Functional Assessment for Burns (FAB), Barthel Index (BI), Timed Up and Go (TUG), Six Minute Walk Test (6MWT), Glasgow Outcome Score (GOS), Rankin Score (RS), Functional Index Repty (FIR), Karnofsky Performance Scale (KPS), Glasgow Coma Score (GCS), Manual Muscle Test (MMT), InterClass Correlation (ICC), Standard Error of the Mean (SEM), Confidence Interval (CI), Seconds (Sec), Minutes (Min), Effect Size Index (ESI), Discharge (DC), Intensive Care Unit (ICU), Odds Ratio (OR), Length of Stay (LOS), Minimally clinically Important Difference (MCID), Minimal Detectable Change at 90% confidence (MDC<sub>90</sub>), Minimal Detectable Difference at a 95% confidence (MDD<sub>95</sub>), Registered Nurse (RN), Physical therapist (PT), Medical Doctor (MD), positive predictive value (PPV), Negative predictive value (NPV), Predictive Value (PV), ICU Acquired Weakness (ICUAW)
Appendix D: Patient Demographic and Clinical Data: Part A

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age/Gender/Race</th>
<th>ICU LOS</th>
<th>Diagnosis</th>
<th>APACHE II Score</th>
<th>SOFA Score</th>
<th>Evaluation physical function measure score</th>
<th>DC from ICU physical function measure score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adm</td>
<td>DC</td>
<td>Adm</td>
<td>DC</td>
<td>Adm</td>
<td>DC</td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key: ICU: Intensive Care Unit. LOS: Length of Stay. Gender: F = Female, M = Male, O = not specified. Race: C = Caucasian/white, B = black or African American, H = Hispanic, A = Asian. APACHE II: 2nd version of the Acute Physiologic Assessment and Chronic Health Evaluation. SOFA: Sequential Organ Failure Assessment. DC: Discharge. Adm = Admission
Appendix D: Patient Demographic and Clinical Data: Part B Inclusion Criteria

<table>
<thead>
<tr>
<th>Patient</th>
<th>MD agreement</th>
<th>Activity Order</th>
<th>RASS Score</th>
<th>PT Consult Date</th>
<th>Vital Signs BP/HR/SpO₂</th>
<th>O₂ needs</th>
<th>I</th>
<th>A</th>
<th>Current Medications</th>
<th>Agree to Study Yes/No</th>
<th>Had to withdrawal from study and why</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix E: Testing Instructions for the 53-Item CPFM

Below are the scales and instructions to be used for the comprehensive physical function assessment.

Instructions for use of the comprehensive physical function assessment:

1. The therapist assessing the patient will verbally explain all tasks to the patient prior to completion.
2. There is no specific order for completion of tasks, for example, if the patient is received in the chair, the patient may complete chair or standing tasks prior to bed mobility tasks.
3. The admission assessment must be started and completed on the same day. The discharge assessment must be started and completed on the same day.
4. The patient will be given an hour to complete as many tasks as possible and may have rest breaks.
5. Assistive devices such as bed railings, walkers, canes, etc. will only be used if indicated.

Richmond Agitation-Sedation Scale

<table>
<thead>
<tr>
<th>Score</th>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4</td>
<td>Combative</td>
<td>Overly combative or violent; immediate danger to staff.</td>
</tr>
<tr>
<td>+3</td>
<td>Very agitated</td>
<td>Pulls on or removes tube(s), catheter(s) or has aggressive behavior towards staff</td>
</tr>
<tr>
<td>+2</td>
<td>Agitated</td>
<td>Frequent non-purposeful movement of patient-ventilator dys-synchrony</td>
</tr>
<tr>
<td>+1</td>
<td>Restless</td>
<td>Anxious or apprehensive but movements are not aggressive or vigorous</td>
</tr>
<tr>
<td>0</td>
<td>Alert and Calm</td>
<td>Not fully alert yet, but has sustained more than 10 seconds awakening, with eye contact, to voice.</td>
</tr>
<tr>
<td>-1</td>
<td>Drowsy</td>
<td>Briefly (less than 10 seconds) awakens with eye contact to voice</td>
</tr>
<tr>
<td>-2</td>
<td>Light sedation</td>
<td>Any movement, but no eye contact to voice</td>
</tr>
<tr>
<td>-3</td>
<td>Moderate sedation</td>
<td>Any movement, but no eye contact to voice</td>
</tr>
</tbody>
</table>
-4  Deep sedation  No response to voice, but any movement to physical stimulation
-5  Unarousable  No response to voice or physical stimulation

- **RASS Procedure:**
  1) Observe patient. Is patient alert and calm? (score 0). Does patient have behavior that is consistent with restlessness or agitation (score +1 to +4 using the criteria listed above.
  2) If patient is not alert, in a loud speaking voice state patient’s name and direct patient to open eye and look at speaker. Repeat once if necessary. Can prompt patient to continue looking at speaker
   a. Patient has eye opening and eye contact, which is sustained for more than 1-seconds (score -1)
   b. Patient has eye opening and eye contact, but cannot sustain for 10 seconds (score -2)
   c. Patient has any movement in response to voice, excluding eye contact (score -3)
  3) If patient does not respond to voice, physically stimulate patient by shaking shoulder and then rubbing sternum if there is no response to shaking shoulder
   a. Patient has any movements to physical stimulation (score -4)
   b. Patient has no response to voice or physical stimulation (score -5)

- **The Acute Physiology and Chronic Health Evaluation (APACHE)**
  - The APACHE II score will be evaluated using an on-line calculator:
    [https://www.mdcalc.com/apache-ii-score](https://www.mdcalc.com/apache-ii-score)

- **The Sequential Organ Failure Assessment (SOFA)**
  - The SOFA score will be evaluated using an on-line calculator:

- **Section I: Strength Testing Instructions**
  - Each muscle group is tested bilaterally. The patient may be given repeat instructions or additional cues if there is a lack of understanding or delayed response. The patient may also be given encouragement if maximum effort is not suspected. Resistance is applied with only one hand and the second hand is used
to stabilize the limb. If a patient has a missing limb, is casted preventing testing, or cannot be placed in the proper position, the patient will receive a score of 0. Proper positioning will vary for each muscle group but should allow for optimum testing against gravity. Gravity eliminated positions should be used to fully differentiate a 2 versus a 1 grade. The best score throughout a session should also be recorded, for example, if tested supine in bed and then tested seated, the highest score should be reflected.

- Grading is as follows:
  - 0 = No visible or palpable evidence of contractility
  - 1 = Visible or palpable evidence of contractility but with no movement
  - 2 = The patient can move through at least partial range of motion
  - 3 = The patient can move the extremity through available range of motion against gravity but cannot hold resistance
  - 4 = The patient can move the extremity through available range of motion against gravity and can hold some resistance
  - 5 = The patient can move the extremity through available range of motion against gravity and can hold with full resistance

➢ Testing Positions
  - **Shoulder Abduction**: The patient should lift their arm out to the side to shoulder level, palms should remain down. Resistance should be applied just over the elbow area (distal humerus).
  - **Shoulder flexion**: The patient should lift their arm straight ahead of them until full active range of motion is evaluated (~180deg). Arm should then be settled at 90deg reaching forward, resistance is applied around elbow area (distal humerus)
  - **Elbow flexion**: The patient is instructed to place their arm by their side (elbow at side) and then to flex their elbow to bring their hand in contact with their shoulder. Once full range has been evaluated, the patient will then be brought to mid-range for resistance to be applied.
  - **Elbow extension**: For optimal scoring of against gravity elbow flexion, the patient’s arm will be elevated overhead and then the patient will be asked to extend their arm towards the ceiling. Upon full completion, the arm will be placed in mid-range and resistance will be applied. In the event that full should range is not avail or arm cannot (for medical reasons) be placed in this position, the patient will be asked to extend arm out after fully flexing the elbow and resistance will be applied at this point.
  - **Wrist Extension**: Arm should be at side, elbow at 90 degrees with arm pronated and the wrist fully extended. Resistance should be applied over the back of the patient’s hand just distal to the wrist.
  - **Ileopsoas/Hip Flexion**: The patient should be positioned so that hips are at 90deg to start and knees flexed. (seated is most optimal). The patient will be cued to bring knee up towards chest. Resistance will be applied on top of the thigh just proximal to the knee.
- **Quadriceps Femoris/Knee Extension**: Seated positioning is again optimal. Instruct patient to fully extend knee. (avoid hyperextension). Once extended, resistance will be applied just proximal to the ankle joint while the second hand supports under the thigh just proximal to the knee.

- **Tibialis Anterior/Ankle Dorsiflexion**: Sitting with the heel on the floor, pull the foot up to full dorsiflexion. Resistance will be given to the dorsum of the foot just proximal to the toes.

- **Grip Strength**: Patients will be assessed for grip strength using a hand-held dynamometer. Correct positioning is arm at side (elbow touching side), elbow at 90deg. Patient’s will be asked to squeeze 3 times. Best score of the three will be recorded. See Tables below.

- **Straight Leg Raise**: Will be assessed from supine. Patient will be considered as having accomplished a straight leg raise if they can achieve 30 degrees of elevation with knee held straight and hold for 5 seconds.

- **Full MRC-SS**: Scores from bilateral shoulder abduction, elbow flexion, wrist extension, hip flexion, knee extension, and dorsiflexion will be totaled and evaluated for the MRC-SS.

- **Age predicted hand grip strengths as noted in table below…**

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Hand</th>
<th>Mean</th>
<th>&lt;20%</th>
<th>&lt;40%</th>
<th>&lt;60%</th>
<th>&gt;60%</th>
<th>&gt;80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 to 19</td>
<td>R</td>
<td>46.91</td>
<td>9.38</td>
<td>18.76</td>
<td>28.15</td>
<td>37.55</td>
<td>37.53</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>42.13</td>
<td>6.43</td>
<td>16.85</td>
<td>25.28</td>
<td>33.70</td>
<td>33.70</td>
</tr>
<tr>
<td>20 to 24</td>
<td>R</td>
<td>48.15</td>
<td>6.63</td>
<td>19.26</td>
<td>28.69</td>
<td>38.52</td>
<td>38.52</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>43.08</td>
<td>6.62</td>
<td>17.23</td>
<td>25.85</td>
<td>34.46</td>
<td>34.46</td>
</tr>
<tr>
<td>25 to 29</td>
<td>R</td>
<td>53.76</td>
<td>10.75</td>
<td>21.50</td>
<td>32.26</td>
<td>43.01</td>
<td>43.01</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>48.60</td>
<td>9.72</td>
<td>19.44</td>
<td>29.16</td>
<td>38.88</td>
<td>38.88</td>
</tr>
<tr>
<td>30 to 34</td>
<td>R</td>
<td>52.63</td>
<td>10.53</td>
<td>21.05</td>
<td>31.58</td>
<td>42.10</td>
<td>42.10</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>48.98</td>
<td>9.80</td>
<td>19.59</td>
<td>29.39</td>
<td>39.18</td>
<td>39.18</td>
</tr>
<tr>
<td>35 to 39</td>
<td>R</td>
<td>53.16</td>
<td>10.63</td>
<td>21.26</td>
<td>31.90</td>
<td>42.53</td>
<td>42.53</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>51.75</td>
<td>10.35</td>
<td>20.70</td>
<td>31.65</td>
<td>41.40</td>
<td>41.40</td>
</tr>
<tr>
<td>40 to 44</td>
<td>R</td>
<td>55.49</td>
<td>11.10</td>
<td>22.20</td>
<td>33.29</td>
<td>44.39</td>
<td>44.39</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>50.40</td>
<td>10.08</td>
<td>20.16</td>
<td>30.24</td>
<td>40.32</td>
<td>40.32</td>
</tr>
<tr>
<td>45 to 49</td>
<td>R</td>
<td>49.93</td>
<td>9.99</td>
<td>19.97</td>
<td>29.96</td>
<td>39.94</td>
<td>39.94</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>46.94</td>
<td>8.79</td>
<td>19.56</td>
<td>29.36</td>
<td>38.15</td>
<td>39.15</td>
</tr>
<tr>
<td>50 to 54</td>
<td>R</td>
<td>48.40</td>
<td>8.68</td>
<td>19.36</td>
<td>29.04</td>
<td>38.72</td>
<td>38.72</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>41.46</td>
<td>8.29</td>
<td>16.58</td>
<td>24.88</td>
<td>33.17</td>
<td>33.17</td>
</tr>
<tr>
<td>55 to 59</td>
<td>R</td>
<td>45.71</td>
<td>9.14</td>
<td>18.26</td>
<td>27.43</td>
<td>36.57</td>
<td>36.57</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>42.16</td>
<td>8.43</td>
<td>16.86</td>
<td>25.30</td>
<td>33.73</td>
<td>33.73</td>
</tr>
<tr>
<td>60 to 64</td>
<td>R</td>
<td>40.59</td>
<td>8.12</td>
<td>16.24</td>
<td>24.35</td>
<td>32.47</td>
<td>32.47</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>37.25</td>
<td>7.45</td>
<td>14.90</td>
<td>22.35</td>
<td>29.80</td>
<td>29.80</td>
</tr>
<tr>
<td>65 to 69</td>
<td>R</td>
<td>40.87</td>
<td>8.17</td>
<td>16.35</td>
<td>24.62</td>
<td>32.70</td>
<td>32.70</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>36.57</td>
<td>7.31</td>
<td>14.63</td>
<td>21.94</td>
<td>29.26</td>
<td>29.26</td>
</tr>
<tr>
<td>70 to 74</td>
<td>R</td>
<td>37.48</td>
<td>7.50</td>
<td>14.99</td>
<td>22.49</td>
<td>29.98</td>
<td>29.98</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>35.49</td>
<td>7.10</td>
<td>14.20</td>
<td>21.29</td>
<td>28.39</td>
<td>28.39</td>
</tr>
<tr>
<td>75+</td>
<td>R</td>
<td>32.76</td>
<td>6.55</td>
<td>13.10</td>
<td>19.66</td>
<td>26.21</td>
<td>26.21</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>28.59</td>
<td>5.72</td>
<td>11.44</td>
<td>17.15</td>
<td>22.87</td>
<td>22.87</td>
</tr>
</tbody>
</table>

➢ **Section II: Bed Mobility**

- The patient will be tested from their hospital bed or stretcher with bed flat and no rails, unless contraindicated for medical reasons. Any use of elevated bed or rails cannot receive an independent score. The highest achieved will be Modified.
Independent (Mod I). Feedback, encouragement, and cues can be given as needed. If unable to perform the task for medical reasons, the patient will be scored a 1.

- **Grading will be as follows:**
  - 1 = Unable to perform the task or requires assist of 2 people to complete
  - 2 = The patient requires Moderate (Mod) assistance (A) defined as the patient doing 26-74% of the work or Maximum (Max) assistance defined as the patient doing < 25% of the work
  - 3 = The patient requires contact guard (CG) defined as touching or guidance but no assistance, or Minimum (Min) A defined as the patient being able to complete 75% or more of the work
  - 4 = The patient requires supervision to complete the task either for cueing purposes or safety
  - 5 = The patient completes the task as Modified Independent (Mod I) defined as use of rails or increased time, or Independent (I) defined as ability to complete the task with the bed flat, no assistance, no devices.

- **Rolling:** The patient will be asked to roll from side to side with the least amount of assistance. A roll will be considered complete if patient can get fully to their lateral hip positioning.

- **Bridging:** The patient will be asked to bend their knees and lift their bottom clear of the bed. The therapist will use their hand as a judge for clearance from the mattress. If the therapist can pass their hands under the patient’s buttock and the patient is able to keep their hands relaxed across their stomach, the patient will be deemed as independent. Modified Independence will be utilized if the patient has to use their arms to push from the mattress, if they use their own arms to support buttock/hips, or if they need to hold railings for leverage. Assist levels will be graded as noted if assist is needed.

- **Supine to sit:** The bed will be placed in full supine (unless medically contraindicated for full supine) Patient’s will be asked to come from supine to sitting with their legs over the edge of the bed. The patient may complete this task at personal preference unless medical condition indicates special techniques (ie log rolling or sternal precautions). If the patient must complete the task in a precautionary manner, the therapist will give instructions first and then allow the patient to proceed. Cueing can be given mid-transfer as needed. Rails should only be used if the patient has rails at home, i.e. baseline hospital bed or installed rails for use at home.

- **Sit to supine:** The bed will be placed in full supine (unless medically contraindicated for full supine). The patient will be asked to go from a seated position to supine in the bed. Again, the patient may complete this task at preference unless precautions are otherwise indicated for safety. Cueing can be given mid-transfer as needed. Rails should only be used if the patient has rails at home, i.e. baseline hospital bed or installed rails for use at home.
Section III: Balance

- The patient will be asked to perform several balance challenges and positions. Cues, encouragement, and instruction can be given throughout testing. Different instructions are noted for each task.
- Grading will be as follows:
  - 1 = Unable to perform the task or requires assist of 2 people to complete
  - 2 = The patient requires Moderate (Mod) assistance (A) defined as the patient doing 26-74% of the work or Maximum (Max) assistance defined as the patient doing < 25% of the work
  - 3 = The patient requires contact guard (CG) defined as touching or guidance but no assistance, or Minimum (Min) A defined as the patient being able to complete 75% or more of the work
  - 4 = The patient requires supervision to complete the task either for cueing purposes or safety
  - 5 = The patient completes the task as Modified Independent (Mod I) defined as use of rails or increased time, or Independent (I) defined as ability to complete the task with the bed flat, no assistance, no devices.
- Static Seated Balance: Patients will be asked to sit edge of bed or at the edge of a chair for at least 10 seconds without holding arm rests, slumping, or swaying. Feet should be resting on the floor if possible. Ability will be graded based on assistance needed over the 10 seconds.
- Dynamic Seated Balance: The patient will be asked to reach outside of their base of support while sitting unsupported. The patient should be asked to reach with dominant hand and in 4-directions: up, down, right, left. Ability will be graded based on the maximum assistance needed with any direction. For example: If the patient requires Min A for Left, right and up but requires Max assist for down, the patient should be graded Max A.
- Static Standing Balance: The patient will be asked if they can stand static in a comfortable stance without an assistive device for 10 seconds. If the patient uses an assistive device baseline, this assistive device may be used for testing, if no assistive device prior to admit, an assistive device will not be used.
- Stand with Feet Together: The patient will then be asked to stand with feet together for 10 seconds. Again, this should be without an assistive device. If the patient uses an assistive device baseline, the assistive device may be used for testing.
- Stand on toes: The patient will be asked to stand on their toes and hold for 10 seconds. The patient can be given assistance for the task via handheld assistance of therapist or by an assistive device. Grading will be as noted.
- Semi-Tandem Stance eye open: The patient will be asked to stand with one foot slightly ahead of the other and hold for 10 seconds. The patient may be given assistance for the task via hand-held assistance of therapist or an assistive device.
- **Tandem Stance with eyes closed:** The patient will be asked to fully place one foot in front of the other and hold for 10 seconds. The patient may be given assistance for the task via hand-held assistance of therapist or an assistive device.
- **Pick up a pen from the floor:** The patient will be asked to pick up a pen from the floor. Assistance will be graded.

### Section IV: Transfers

- Transfers should be performed from a “reasonable” surface. Optimal starting position would be with hips at 90deg. Patient’s may be cued, encouraged, or given reminders during the task. Transfers should not be assessed with bed rail usage unless patient has access to bed rails at home. If the patient performs the transfer from several surfaces during a session, the lowest score should be recorded. For example, sit to stand from the bed is CG, sit to stand from a chair is Min A, but from the toilet is Mod A. The Mod A should be recorded.
- Grading will be as follows:
  - 1 = Unable to perform the task or requires assist of 2 people to complete
  - 2 = The patient requires Moderate (Mod) assistance (A) defined as the patient doing 26-74% of the work or Maximum (Max) assistance defined as the patient doing < 25% of the work
  - 3 = The patient requires contact guard (CG) defined as touching or guidance but no assistance, or Minimum (Min) A defined as the patient being able to complete 75% or more of the work
  - 4 = The patient requires supervision to complete the task either for cueing purposes or safety
  - 5 = The patient completes the task as Modified Independent (Mod I) defined as use of rails or increased time, or Independent (I) defined as ability to complete the task with the bed flat, no assistance, no devices.
- **Sit to stand/Stand to sit:** Performed preferably from a 90 deg hip flexion position, may use chair arms or commode arms, should not use bed rails unless utilized at home.
- **Bed to chair/Chair to bed:** The surfaces should be somewhat equal in height to perform the transfer. Many ICU and hospital settings have very high bed surfaces which are unrealistic for many patients. In this case, transfers should be tested chair to chair/commode/WC; to allow for accurate scoring.
- **Stand without the use of arms:** Patient should be sitting with hips preferably at a 90 deg angle and will be asked to stand without use of arms.

### Section V: Gait

- Gait should be tested on an even surface. Avoid inclines. Patient’s may be given cues, encouragement, and feedback during ambulation. If a patient utilized an assistive device prior to admit, the patient may use their assistive device. If they did not use an assistive device baseline, they may only use a device where noted. Grading for each task will be as noted below.
- **Step Cadence**: Patient’s will be advised to stand and march in place for once minute taking as many steps as they can. Patients can be supported by hand-held assist of the therapist or by an assist device is needed baseline. Patients will then be given a seated break. Steps will be scored as noted: 1) unable, 2) 1-49 steps/min, 3) 50-79 steps/min, 4) 80-99 steps/min, and 5) > 100 steps/min.

- **Ambulation**: will be scored on assistance level with and without a device. Patient’s will be scored as follows 1) unable to complete or assist of 2 needed, 2) Mod/Max Assist, 3) CG/Min A, 4) supervision, or 5) Mod I / I

- **Ability to walk backwards**: Patient’s will be cued to take 4 steps backwards and will be graded on level of assistance needed for this task as follows: 1) unable to complete or assist of 2 needed, 2) Mod/Max Assist, 3) CG/Min A, 4) supervision, or 5) Mod I / I

- **Distance and Time walked**: Patient’s will be advised to walk for as long as they can. Patient’s may take standing breaks as needed. The therapist will monitor distance as well as overall time until a seated break is needed. The grading will be as follows: Distance 1) 0-25 feet, 2) 26-50 feet, 3) 51-99 feet, 4) 100-199 feet, or 5) ≥ 200’ and Time 1) unable, 2) < 2 minutes, 3) 2-4 minutes, 4) 4-5 minutes, 5) > 5 minutes.

- **Jump**: Patient’s will be asked to jump. The patient will be supported by the therapist for any assistance needed but may not push from a walker or hold rails.

➢ **Section VI: Stairs**

- **Stairs** will be tested based on numbers of steps completed and assistance needed. Patient’s may use railings or assistive devices as they would at home. If no railings are available at home, no railings should be used for testing. The patient may ascend/descend the steps however comfortable, for example: step-over-step, or step-to-step.

- **Ascend/descend steps**: The patient will be advised to climb as many steps as they can

- **Stair Assistance**: The patient will be scored as follows for assistance level.
  - 1 = Unable to perform the task or requires assist of 2 people to complete
  - 2 = The patient requires Moderate (Mod) assistance (A) defined as the patient doing 26-74% of the work or Maximum (Max) assistance defined as the patient doing < 25% of the work
  - 3 = The patient requires contact guard (CG) defined as touching or guidance but no assistance, or Minimum (Min) A defined as the patient being able to complete 75% or more of the work
  - 4 = The patient requires supervision to complete the task either for cueing purposes or safety
  - 5 = The patient completes the task as Modified Independent (Mod I) defined as use of rails or increased time, or Independent (I) defined as ability to complete the task with no assistance or devices.
Section VII: Medical Status

- This section is designed to reflect medical complexity of the patient. This is to be scored at the time of the session.

- Respiratory function: This section will indicate oxygen needs. If the patient requires mechanical ventilation, they will be scored a 1. If the patient requires non-invasive ventilation such as Bi-pap or high flow Nasal cannula machines (ie. Vapotherm or Oxiflow), they will be scored a 2. Should the patient require ≥ 6 liters of oxygen or the equivalent of ≥ 48% FiO2 via supplemental oxygen, they will score a 3. If the patient requires < 6 liters per minute or the equivalent of < 48% FiO2 via supplemental oxygen, they will score a 4. If the patient is stable on room air, they will score a 5.
  - Note: if patient is received on oxygen and is trialed on room air for the entire session and tolerates it, the patient will be marked as room air. If the patient is trialed on room air but requires return to oxygen, the max amount of oxygen needed will be recorded. If the patient needs increased oxygen needs from baseline oxygen for activity of ambulation, the max amount of oxygen needed will be marked. For example, if the patient require return to Bipap post session from a nasal cannula, the Bipap will be marked as the level of oxygen.

- Cough Ability: This section will rate the quality of a patient’s cough. If the patient is unable to illicit a cough, they will score a 1. If the cough can be stimulated with suctioning or a tongue depressor, they will score a 2. If the patient’s cough is weak but still requires assistance with suctioning for full clearance, they will score a 3. If the patient’s cough is weak but they can clear their airway or suction self as needed, they will score a 4. If the patient has a consistent, volitional cough, they will score a 5.

- Mental Alertness/Command Following: The patient will be graded on their ability to follow commands. The following percentages will indicate their level. 1) unable to follow commands, 2) follows < 25% of commands, 3) follows 26-50% of commands, 4) follows 51-75% of commands, 5) follows ≥76% of commands.
  - The patient will be asked to complete 4 tasks. Completion of these tasks will be graded for percentage of following. The tasks will be:
    - Open your eyes
    - Stick out your tongue
    - Show me “thumbs up”
    - Show me “two fingers”

- Pain: The level of pain reported by the patient will be indicated. If pain varies through the session, the max pain reported will be indicated.
  - The Numeric Rating Scale Instructions “Please indicate the intensity of current pain level a scale of 0 (no pain) to 10 (worst pain imaginable)”
Mobility Barriers: indication to the number of following lines or tubes: supplemental oxygen, foley catheter, endotracheal tube (ETT), tracheostomy (Trach), central line, peripheral intravenous line (IV), arterial line, dialysis catheter, peripherally inserted central catheter (PICC), PEG/PEJ tube, nasogastric tube (NGT), chest tube (CT), temporary pacemaker, pulmonary artery catheter, patient-controlled analgesia (PCA) or epidural, Intra-aortic balloon pump (IABP), Continuous dialysis, ventriculostomy or eternal ventricular drain (EVD), wound vac, or other.

Continuous intravenous drips that cannot be disconnected for mobilization. If yes, the patient will score 1. If no, the patient will score a 5.

Does the patient currently have any medical contra-indications for OOB therapy?
Appendix F: Item Fit Decision Characteristics for the 53-Item CPFM

Wright Map Admission Data

Key: stairass = stair assist, penfromf = pen from the floor, semi-tan = semi-tandem stance, stepcade = step cadence, gripL = grip left, standont = stand on toes, ambwdevi = ambulate with a device, ambwdev = ambulate without a device, bed_chai = bed to chair, chair_be = chair to bed, feettoge = stand with feet together, gripR = grip right, sit_stan = sit to stand, standnoa = stand no arms, stand_si = stand to sit, sit_supi = sit to supine, standbal = static standing balance, sup_sit = supine to sit, ivgtts_a = intravenous drips on admission, rollL = roll left, rollR = roll R, dynseab = dynamic seated balance, lines_AD = lines present on admission, pain_adm = pain on admission, seatbala = static seated balance, SLR = straight leg raise, resp_func = respiratory function on admission, deltoidL = deltoid left, MRCSS = medical research council sum score, V4-A = shoulder flexion left, bicepL = bicep left, cough_AD = cough on admission, deltoid = deltoid R, DFL = dorsiflexion left, hipflexL = hip flexion left, hipflexR = hip flexion right, quadL = quadriceps left, shldflex = shoulder flexion, tricepL = tricep left, tricepR = tricep right, V10_A = wrist extension left, alert_AD, mental alertness on admission, bicepR = bicep right, DFR = dorsiflexion right, quadR = quadriceps right, wristtext = wrist extension R
# Wright Map Discharge Data

<table>
<thead>
<tr>
<th>MEASURE</th>
<th>PERSON - MAP - ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>X T+</td>
</tr>
<tr>
<td></td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>XX</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>XXX</td>
</tr>
<tr>
<td></td>
<td>XXXXXX</td>
</tr>
<tr>
<td>4</td>
<td>XXXXXX S+</td>
</tr>
<tr>
<td></td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td>JUMP</td>
</tr>
<tr>
<td></td>
<td>XXXX XXXX</td>
</tr>
<tr>
<td></td>
<td>TANDEM</td>
</tr>
<tr>
<td>3</td>
<td>XXXXXXX</td>
</tr>
<tr>
<td></td>
<td>GRIPL</td>
</tr>
<tr>
<td></td>
<td>PENFROMF</td>
</tr>
<tr>
<td></td>
<td>STEPCODE</td>
</tr>
<tr>
<td></td>
<td>XXXXXXX</td>
</tr>
<tr>
<td></td>
<td>GRIPR</td>
</tr>
<tr>
<td></td>
<td>STAIRSS</td>
</tr>
<tr>
<td></td>
<td>SEMI_TAN</td>
</tr>
<tr>
<td></td>
<td>STAIRASS</td>
</tr>
<tr>
<td></td>
<td>STANDONT</td>
</tr>
<tr>
<td>2</td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td></td>
<td>=S TIME</td>
</tr>
<tr>
<td></td>
<td>XXXXXXXXX</td>
</tr>
<tr>
<td></td>
<td>AMBDEV</td>
</tr>
<tr>
<td></td>
<td>BACKWARD</td>
</tr>
<tr>
<td></td>
<td>DISTANCE</td>
</tr>
<tr>
<td></td>
<td>XXXXXXX</td>
</tr>
<tr>
<td></td>
<td>BED_CHAI</td>
</tr>
<tr>
<td></td>
<td>CHAIR_BE</td>
</tr>
<tr>
<td></td>
<td>SUP_SIT</td>
</tr>
<tr>
<td>1</td>
<td>XX + SIT_STAN</td>
</tr>
<tr>
<td></td>
<td>SIT_SUPI</td>
</tr>
<tr>
<td></td>
<td>STANDBAL</td>
</tr>
<tr>
<td></td>
<td>STAND_SI</td>
</tr>
<tr>
<td></td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td>XXXX S</td>
</tr>
<tr>
<td></td>
<td>ROLL</td>
</tr>
<tr>
<td></td>
<td>ROLLR</td>
</tr>
<tr>
<td>0</td>
<td>XX + S</td>
</tr>
<tr>
<td></td>
<td>BRIDGE</td>
</tr>
<tr>
<td></td>
<td>XXXX</td>
</tr>
<tr>
<td></td>
<td>PAIN_DC</td>
</tr>
<tr>
<td></td>
<td>SEATBALA</td>
</tr>
<tr>
<td></td>
<td>SLR</td>
</tr>
<tr>
<td>-1</td>
<td>X + IVGTTs_D</td>
</tr>
<tr>
<td></td>
<td>LINES_DC</td>
</tr>
<tr>
<td></td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>XX T</td>
</tr>
<tr>
<td></td>
<td>MRCSS</td>
</tr>
<tr>
<td></td>
<td>RESP_FUNC</td>
</tr>
<tr>
<td></td>
<td>V10_A</td>
</tr>
<tr>
<td></td>
<td>V4_A</td>
</tr>
<tr>
<td>-2</td>
<td>+S DELTOIDL</td>
</tr>
<tr>
<td></td>
<td>HIPFLEXXR</td>
</tr>
<tr>
<td></td>
<td>TRICEPL</td>
</tr>
<tr>
<td></td>
<td>SHLDFLEX</td>
</tr>
<tr>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>BICEPL</td>
</tr>
<tr>
<td></td>
<td>COUGH_DC</td>
</tr>
<tr>
<td></td>
<td>TRICEPR</td>
</tr>
<tr>
<td></td>
<td>WRISTEXT</td>
</tr>
<tr>
<td>-3</td>
<td>X +</td>
</tr>
<tr>
<td></td>
<td>BICEPR</td>
</tr>
<tr>
<td></td>
<td>ALERT_DC</td>
</tr>
<tr>
<td>-4</td>
<td>&lt;less&gt;</td>
</tr>
<tr>
<td></td>
<td>&lt;freq&gt;</td>
</tr>
</tbody>
</table>

Key: stairass = stair assist, penfromf = pen from the floor, semi-tan = semi-tandum stance, stepcada = step cadence, gripL = grip left, standont = stand on toes, ambwdevi = ambulate with a device, ambwodevi = ambulate without a device, bed_chai = bed to chair, chair_be = chair to bed, feettoe = stand with feet together, gripR = grip right, sit站长 = sit to stand, standnoa = stand no arms, stand_si = stand to sit, sit_supi = sit to supine, standbal = static standing balance, sup_sit = supine to sit, ivgts_a = intravenous drips on admission, rollL = roll left, rollR = roll R, dynseatb = dynamic seated balance, lines_AD = lines present on admission, pain_adm = pain on admission, seatbala = static seated balance, SLR = straight leg raise, resp_func = respiratory function on admission, deltoidL = deltoid left, MRCSs = medical research council sum score, V4-A = shoulder flexion left, bicepL = bicep left, cough_AD = cough on admission, deltold = deltoid R, DFL = dorsiflexion left, hipflexL = hip flexion left, hipflexR = hip flexion right, quadL = quadriceps left, shldflex = shoulder flexion, tricepL = triceps left, tricepR = triceps right, V10_A = wrist extension left, alert_AD, mental alertness on admission, bicepR = bicep right, DFR = dorsiflexion right, quadR = quadriceps right, wristext = wrist extension R
### Strength and Hand Grip Comparison

<table>
<thead>
<tr>
<th></th>
<th>Out-Fit (&gt; 2.0 logits)</th>
<th>Dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original Comprehensive Physical Function Measure Likert Scale for Strength and Hand Grip</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wrist ext DC</td>
<td>2.56 (.57)</td>
<td>Raw Variance: 47.3%</td>
</tr>
<tr>
<td>DF right</td>
<td>3.05 (.46)</td>
<td>Unexplained 1&lt;sup&gt;st&lt;/sup&gt; Variance: 7.11, 2.4%</td>
</tr>
<tr>
<td>DF left</td>
<td>2.37 (.49)</td>
<td>Unexplained 2&lt;sup&gt;nd&lt;/sup&gt; Variance: 5.48, 1.8%</td>
</tr>
<tr>
<td>Grip right</td>
<td>2.49 (.79)</td>
<td></td>
</tr>
<tr>
<td>Resp func DC</td>
<td>2.10 (.51)</td>
<td></td>
</tr>
<tr>
<td>Pain admit</td>
<td>7.37 (.70)</td>
<td></td>
</tr>
<tr>
<td>Pain DC</td>
<td>3.58 (.61)</td>
<td></td>
</tr>
<tr>
<td>IVGtts Admit</td>
<td>9.90 (.72)</td>
<td></td>
</tr>
<tr>
<td>IVGtts DC</td>
<td>9.70 (.58)</td>
<td></td>
</tr>
<tr>
<td><strong>Differential Coding for Weak/Strong and use of Age Predicted Hand Grip Strength</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grip Strong</td>
<td>2.68 (.73)</td>
<td>Raw Variance: 46.1%</td>
</tr>
<tr>
<td>Grip Weak</td>
<td>2.72 (.71)</td>
<td>Unexplained 1&lt;sup&gt;st&lt;/sup&gt; Variance: 8.23, 2.9%</td>
</tr>
<tr>
<td>RespFunc DC</td>
<td>2.23 (.54)</td>
<td>Unexplained 2&lt;sup&gt;nd&lt;/sup&gt; Variance: 5.48, 1.9%</td>
</tr>
<tr>
<td>Pain Admit</td>
<td>3.83 (.69)</td>
<td></td>
</tr>
<tr>
<td>Pain DC</td>
<td>3.41 (.62)</td>
<td></td>
</tr>
<tr>
<td>IVGtts Admit</td>
<td>6.26 (.72)</td>
<td></td>
</tr>
<tr>
<td>IVGtts DC</td>
<td>7.74 (.59)</td>
<td></td>
</tr>
</tbody>
</table>

Key: ext = extension, DC = discharge, resp func = respiratory function, IVGtts = intravenous drips,
## Appendix G: Initial 53-Item Item Order and Outfit Data

### Complete Table of Item Order Entry

<table>
<thead>
<tr>
<th>Item</th>
<th>Logit Measure</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deltoid R</td>
<td>-2.77</td>
<td>0.13</td>
</tr>
<tr>
<td>Deltoid L</td>
<td>-2.19</td>
<td>0.12</td>
</tr>
<tr>
<td>Shoulder Flexion R</td>
<td>-2.33</td>
<td>0.13</td>
</tr>
<tr>
<td>Shoulder Flexion L (V4_A)</td>
<td>-1.81</td>
<td>0.11</td>
</tr>
<tr>
<td>Bicep R</td>
<td>-3.13</td>
<td>0.16</td>
</tr>
<tr>
<td>Bicep L</td>
<td>-2.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Tricep R</td>
<td>-2.62</td>
<td>0.13</td>
</tr>
<tr>
<td>Tricep L</td>
<td>-2.18</td>
<td>0.12</td>
</tr>
<tr>
<td>Wrist Extension R</td>
<td>-2.61</td>
<td>0.14</td>
</tr>
<tr>
<td>Wrist Extension L (V10_A)</td>
<td>-1.86</td>
<td>0.11</td>
</tr>
<tr>
<td>Hip Flexion R</td>
<td>-2.02</td>
<td>0.12</td>
</tr>
<tr>
<td>Hip Flexion L</td>
<td>-2.50</td>
<td>0.12</td>
</tr>
<tr>
<td>Quad R</td>
<td>-2.93</td>
<td>0.15</td>
</tr>
<tr>
<td>Quad L</td>
<td>-2.45</td>
<td>0.13</td>
</tr>
<tr>
<td>Dorsiflexion R</td>
<td>-2.70</td>
<td>0.14</td>
</tr>
<tr>
<td>Dorsiflexion L</td>
<td>-1.81</td>
<td>0.12</td>
</tr>
<tr>
<td>Grip Strength R</td>
<td>0.89</td>
<td>0.09</td>
</tr>
<tr>
<td>Grip Strength L</td>
<td>0.79</td>
<td>0.09</td>
</tr>
<tr>
<td>Straight Leg Raise</td>
<td>-0.02</td>
<td>0.08</td>
</tr>
<tr>
<td>MRC-SS</td>
<td>-1.30</td>
<td>0.10</td>
</tr>
<tr>
<td>Rolling R</td>
<td>0.66</td>
<td>0.08</td>
</tr>
<tr>
<td>Rolling L</td>
<td>0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>Ability to Bridge</td>
<td>0.32</td>
<td>0.08</td>
</tr>
<tr>
<td>Supine to Sit</td>
<td>1.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Sit to Supine</td>
<td>1.17</td>
<td>0.08</td>
</tr>
<tr>
<td>Seated Balance</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>Dynamic Seated Balance</td>
<td>0.18</td>
<td>0.08</td>
</tr>
<tr>
<td>Standing Balance</td>
<td>1.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Stand with Feet Together</td>
<td>1.71</td>
<td>0.07</td>
</tr>
<tr>
<td>Standing on Toes</td>
<td>2.64</td>
<td>0.08</td>
</tr>
<tr>
<td>Semi-Tandem Stance</td>
<td>2.69</td>
<td>0.08</td>
</tr>
<tr>
<td>Tandem Stance</td>
<td>3.46</td>
<td>0.08</td>
</tr>
<tr>
<td>Pick up Pen from Floor</td>
<td>2.99</td>
<td>0.08</td>
</tr>
<tr>
<td>Sit to Stand</td>
<td>1.41</td>
<td>0.08</td>
</tr>
<tr>
<td>Stand to Sit</td>
<td>1.41</td>
<td>0.08</td>
</tr>
<tr>
<td>Bed to Chair</td>
<td>1.53</td>
<td>0.07</td>
</tr>
<tr>
<td>Chair to Bed</td>
<td>1.53</td>
<td>0.07</td>
</tr>
<tr>
<td>Stand without Arms</td>
<td>1.81</td>
<td>0.07</td>
</tr>
<tr>
<td>Step Cadence</td>
<td>2.99</td>
<td>0.09</td>
</tr>
<tr>
<td>Item</td>
<td>Logit Measure</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Ambulate without a Device</td>
<td>1.87</td>
<td>0.08</td>
</tr>
<tr>
<td>Ambulate with a Device</td>
<td>1.70</td>
<td>0.07</td>
</tr>
<tr>
<td>Walk Backwards</td>
<td>1.99</td>
<td>0.08</td>
</tr>
<tr>
<td>Distance Ambulated</td>
<td>2.08</td>
<td>0.07</td>
</tr>
<tr>
<td>Time Ambulated</td>
<td>0.42</td>
<td>0.09</td>
</tr>
<tr>
<td>Jump</td>
<td>2.29</td>
<td>0.10</td>
</tr>
<tr>
<td>Number of Stairs</td>
<td>3.11</td>
<td>0.08</td>
</tr>
<tr>
<td>Stair Assistance</td>
<td>3.04</td>
<td>0.08</td>
</tr>
<tr>
<td>Respiratory Function</td>
<td>-1.49</td>
<td>0.10</td>
</tr>
<tr>
<td>Coughing Ability</td>
<td>-2.65</td>
<td>0.13</td>
</tr>
<tr>
<td>Mental Alertness</td>
<td>-2.98</td>
<td>0.18</td>
</tr>
<tr>
<td>Presence of Pain</td>
<td>-0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>Presence of Lines</td>
<td>-0.89</td>
<td>0.09</td>
</tr>
<tr>
<td>Presence of IV Drips</td>
<td>0.42</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Key: R = right, L = left, IV = intravenous  
Blue Highlight: Four easiest item to complete  
Green Highlight: Five hardest items to complete

---

**Complete Table of Outfit Values**

<table>
<thead>
<tr>
<th>Item</th>
<th>Outfit MNSQ (ZTSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deltoid R</td>
<td>0.60 (-1.37)</td>
</tr>
<tr>
<td>Deltoid L</td>
<td>0.51 (-1.80)</td>
</tr>
<tr>
<td>Shoulder Flexion R</td>
<td>0.52 (-1.58)</td>
</tr>
<tr>
<td>Shoulder Flexion L (V4_A)</td>
<td>0.45 (-1.97)</td>
</tr>
<tr>
<td>Bicep R</td>
<td>0.54 (-1.20)</td>
</tr>
<tr>
<td>Bicep L</td>
<td>0.52 (-1.32)</td>
</tr>
<tr>
<td>Tricep R</td>
<td>0.52 (-1.33)</td>
</tr>
<tr>
<td>Tricep L</td>
<td>0.80 (-0.46)</td>
</tr>
<tr>
<td>Wrist Extension R</td>
<td>1.32 (.082)</td>
</tr>
<tr>
<td>Wrist Extension L (V10_A)</td>
<td>1.78 (1.82)</td>
</tr>
<tr>
<td>Hip Flexion R</td>
<td>0.62 (-1.04)</td>
</tr>
<tr>
<td>Hip Flexion L</td>
<td>1.10 (0.40)</td>
</tr>
<tr>
<td>Quad R</td>
<td>0.39 (-1.65)</td>
</tr>
<tr>
<td>Quad L</td>
<td>0.53 (-1.31)</td>
</tr>
<tr>
<td><strong>Dorsiflexion R</strong></td>
<td><strong>2.08 (2.07)</strong></td>
</tr>
<tr>
<td>Dorsiflexion L</td>
<td>1.35 (0.93)</td>
</tr>
<tr>
<td><strong>Grip Strength R</strong></td>
<td><strong>2.18 (9.90)</strong></td>
</tr>
<tr>
<td>Grip Strength L</td>
<td>1.76 (7.21)</td>
</tr>
<tr>
<td>Straight Leg Raise</td>
<td>1.20 (2.72)</td>
</tr>
<tr>
<td>MRC-SS</td>
<td>0.71 (-1.02)</td>
</tr>
<tr>
<td>Rolling R</td>
<td>0.88 (-1.23)</td>
</tr>
<tr>
<td>Activity</td>
<td>Value</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Rolling L</td>
<td>0.93 (-0.65)</td>
</tr>
<tr>
<td>Ability to Bridge</td>
<td>0.73 (-1.88)</td>
</tr>
<tr>
<td>Supine to Bridge</td>
<td>0.98 (-0.09)</td>
</tr>
<tr>
<td>Sit to Supine</td>
<td>1.00 (0.00)</td>
</tr>
<tr>
<td>Seated Balance</td>
<td>0.99 (0.04)</td>
</tr>
<tr>
<td>Dynamic Seated Balance</td>
<td>0.41 (-7.13)</td>
</tr>
<tr>
<td>Standing Balance</td>
<td>0.40 (-7.81)</td>
</tr>
<tr>
<td>Stand with Feet Together</td>
<td>0.45 (-6.37)</td>
</tr>
<tr>
<td>Standing on Toes</td>
<td>0.82 (-1.57)</td>
</tr>
<tr>
<td>Semi-Tandem Stance</td>
<td>0.52 (-4.24)</td>
</tr>
<tr>
<td>Tandem Stance</td>
<td>0.55 (-3.40)</td>
</tr>
<tr>
<td>Pick up Pen from Floor</td>
<td>0.41 (-3.13)</td>
</tr>
<tr>
<td>Sit to Stand</td>
<td>0.34 (-8.36)</td>
</tr>
<tr>
<td>Stand to Sit</td>
<td>0.31 (-8.44)</td>
</tr>
<tr>
<td>Bed to Chair</td>
<td>0.30 (-8.82)</td>
</tr>
<tr>
<td>Chair to Bed</td>
<td>0.27 (-8.69)</td>
</tr>
<tr>
<td>Stand without Arms</td>
<td>0.43 (-5.91)</td>
</tr>
<tr>
<td>Step Cadence</td>
<td>0.72 (-3.25)</td>
</tr>
<tr>
<td>Ambulate without a Device</td>
<td>0.37 (-7.60)</td>
</tr>
<tr>
<td>Ambulate with a Device</td>
<td>0.37 (-7.33)</td>
</tr>
<tr>
<td>Walk Backwards</td>
<td>0.38 (-7.41)</td>
</tr>
<tr>
<td>Distance Ambulated</td>
<td>0.68 (-1.94)</td>
</tr>
<tr>
<td>Time Ambulated</td>
<td>0.75 (-3.32)</td>
</tr>
<tr>
<td>Jump</td>
<td>0.94 (-0.51)</td>
</tr>
<tr>
<td>Number of Stairs</td>
<td>0.45 (-1.80)</td>
</tr>
<tr>
<td>Stair Assistance</td>
<td>0.42 (-2.07)</td>
</tr>
<tr>
<td>Respiratory Function</td>
<td>2.27 (4.35)</td>
</tr>
<tr>
<td>Coughing Ability</td>
<td>0.83 (-0.48)</td>
</tr>
<tr>
<td>Mental Alertness</td>
<td>0.65 (-0.86)</td>
</tr>
<tr>
<td>Presence of Pain</td>
<td>9.90 (9.91)</td>
</tr>
<tr>
<td>Presence of Lines</td>
<td>1.75 (4.51)</td>
</tr>
<tr>
<td>Presence of IV Drips</td>
<td>9.90 (9.91)</td>
</tr>
</tbody>
</table>

Key: R = right, L = left, IV = intravenous
Blue Highlight: Five Outfit Items
Appendix H: Probability Curves for the 15 Individual Items of the CPFM

Probability Curve of the MRC-SS

Probability Curve of Rolling to the Right
Probability Curve of Rolling to the Left

Probability Curve for Bridging
Probability Curve for Supine to Sit

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Expected</th>
<th>Hit</th>
<th>Miss</th>
<th>Andrich</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
</tbody>
</table>

Summary of Category Structure: Model = “X”
FOR GROUPING “5” Item Number: 5 SUP_SIT
Category: 5
Probability Measure of -9.89

Probability Curve for Mental Alertness

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Expected</th>
<th>Hit</th>
<th>Miss</th>
<th>Andrich</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>22</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0.350</td>
<td>2</td>
</tr>
</tbody>
</table>

Summary of Category Structure: Model = “X”
FOR GROUPING “6” Item Number: 6 MENTAL
Category: 6
Probability Measure of -5.78
### Probability Curve for Seated Balance

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Score Mean</th>
<th>Average Expect</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>Threshold</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.6</td>
<td>-4.46</td>
<td>0.46</td>
<td>0.36</td>
<td>-5.36</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>34.3</td>
<td>-2.68</td>
<td>0.82</td>
<td>0.33</td>
<td>-3.13</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>48.9</td>
<td>-2.04</td>
<td>0.63</td>
<td>0.44</td>
<td>-0.91</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>59.3</td>
<td>1.73</td>
<td>0.68</td>
<td>0.67</td>
<td>4.93</td>
<td>5</td>
</tr>
</tbody>
</table>

### Probability Curve for Standing on Toes

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Score Mean</th>
<th>Average Expect</th>
<th>MNSQ</th>
<th>MNSQ</th>
<th>Threshold</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.6</td>
<td>-4.46</td>
<td>0.46</td>
<td>0.36</td>
<td>-5.36</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>34.3</td>
<td>-2.68</td>
<td>0.82</td>
<td>0.33</td>
<td>-3.13</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>48.9</td>
<td>-2.04</td>
<td>0.63</td>
<td>0.44</td>
<td>-0.91</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>59.3</td>
<td>1.73</td>
<td>0.68</td>
<td>0.67</td>
<td>4.93</td>
<td>5</td>
</tr>
</tbody>
</table>
Probability Curve for Picking up a Pen

Probability Curve for Sit to Stand
### Probability Curve for Chair to Bed

**Summary of Category Structure, Model=R**
For grouping "A" Item number: 11 CHAIR
e
**Item Difficulty Measure of .50 Added to Measures**

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Expected</th>
<th>Threshold</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

### Probability Curve for Ambulation without a Device

**Summary of Category Structure, Model=R**
For grouping "C" Item number: 12 AMBMOD

**Item Difficulty Measure of 1.50 Added to Measures**

<table>
<thead>
<tr>
<th>Category</th>
<th>Observed</th>
<th>Expected</th>
<th>Threshold</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

200
Probability Curve for Time of Ambulation

Probability Curve for Ability to Jump
Probability Curve for Completion of Stairs

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>OBSERVED</th>
<th>EXPECTED</th>
<th>Z-Score</th>
<th>THRESHOLD</th>
<th>MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>5</td>
<td>-0.67</td>
<td>1.81</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>6</td>
<td>1.29</td>
<td>1.39</td>
<td>.99</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>8</td>
<td>1.28</td>
<td>2.82</td>
<td>.96</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>18</td>
<td>3.44</td>
<td>4.39</td>
<td>1.36</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>32</td>
<td>6.00</td>
<td>5.74</td>
<td>.42</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>10</td>
<td>6.00</td>
<td>5.74</td>
<td>.42</td>
</tr>
</tbody>
</table>

CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections

Person [MINUS] Item MEASURE

202
Appendix I: Letter of Approval from the Our Lady of Lourdes IRB

February 26, 2019

Michelle L. Peterson, PT, DPT, PhD, CCS
Lourdes Health System
1601 Haddon Avenue
Camden, NJ 08103

RE: Protocol Title: Creation of a Comprehensive Physical Function Measure for the Intensive Care Unit through the Use of Item Response Theory
Principal Investigator: Michelle L. Peterson, PT, DPT, PhD, CCS
Sub-Investigators: Laura Fous, Gina Ragonese
IRB # 19-005
Protocol # None
IRB Approval: February 21, 2019
Expiration Date: February 20, 2020

Dear Dr. Peterson:

The Chair of the Institutional Review Board has reviewed the submitted application materials. The information that you have provided indicates that the study poses no more than minimal risk to subjects and that the research involves only procedures listed in Expedited Review category. The IRB Chair has approved your request for Expedited Review. This study will be reviewed at the next meeting; you will be notified immediately if there is a concern.

Subjects: The number of subjects in this chart review is 250.

This study will serve to create a comprehensive physical function measure for the Intensive Care Unit through the use of Item Response Theory. The goal is to capture all levels of function from bed level to ambulatory, dependent to independent, and to limit the need for multiple outcomes measures to be used with a patient in the ICU, regardless of acuity and/or diagnosis. This will serve to accurately and efficiently determine progression or the lack of progression over time in a quantifiable manner can improve patient outcomes and reduce cost of care.

Amendments/Modifications/Revisions: You are required to seek the approval of the IRB for any amendments or revisions to the protocol prior to implementation except when necessary to eliminate immediate hazards to the subjects. All protocol revisions must be submitted (original plus a copy) with the revisions highlighted along with a copy of the previously approved document.

Continuing Review: Approval is valid until the protocol expiration date shown above. The IRB must review and approve all human subject research studies at intervals appropriate to the degree of risk, but not less than once a year. In order to avoid lags in approval of your research and the suspension of study enrollment, submit your continuing review application at least eight weeks before the study expiration date. Missed submissions are the responsibility of the Principal Investigator regardless of whether or not the IRB notifies you. The continuation of research after expiration of IRB approval is a violation of the regulations governing research.

Completion of Study: Notify the IRB when your study has been stopped for any reason. Submit a written notification by completing the Continuing Review/Request for Termination report form. Neither study closure by the sponsor or the investigator removes the obligation for timely continuing review or a final report. If your study ends prior to approval expiration, or you leave the institution prior to study closure, it is your obligation to notify the Board and submit a final progress report to the Board.

Thank you for your interest in this valuable research project.

Sincerely,

John S. Radziunski, MD
Chair, Institutional Review Board
Appendix J: Our Lady of Lourdes IRB Extension
Appendix K: Site Approval Letter

SITE APPROVAL LETTER

Nova Southeastern University
3301 College Avenue
Fort Lauderdale, FL 33314-7796

Subject: Site Approval Letter

To Whom It May Concern:

This letter acknowledges that I have received and reviewed a request by Michelle Peterson to conduct a research project entitled "Creation of a Comprehensive Physical Function Measure for the Intensive Care Unit through the use of Item Response Theory" and I approve of this research to be conducted at our facility.

When the researcher receives approval for this research project from the Nova Southeastern University's Institutional Review Board/NSU IRB, I agree to provide access for the approved research project. If we have any concerns or need additional information, we will contact the Nova Southeastern University's IRB at (954) 262-5369 or irb@nova.edu.

Sincerely,

Stacey Jarrell, MPT
Director of Therapy
(856)757-3792
jarrells@lourdesnet.org
References

41. Jette AM. Using health-related quality of life measures in physical therapy outcomes research. 


170. Palmieri J, Orest MR. Improving the care of patients who have difficulty weaning from the ventilator in the acute care setting. *Journal of Acute Care Physical Therapy.* 2012;3(2):193-203.
