DEVELOPMENT OF CROSSWALKS TO AGGREGATE INTERNATIONAL SPINAL CORD INJURY FUNCTIONAL DATA

by

LINDA ANN TRUETT JONES-NORSE

B.S., University of Colorado, Boulder, Colorado, 1988

B.S., University of Colorado, Denver, Colorado, 1990

M.S., Samuel Merritt College, Oakland, California, 2001

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Clinical Science

2020

This thesis for the Doctor of Philosophy degree in Clinical Science by

Linda Ann Truett Jones -Norse

has been approved for the

Clinical Science Graduate Program

by

Heather Haugen, Chair

David Weitzenkamp, Advisor

Gale Whiteneck, Mentor

John Steeves

Susan Charlifue

Date: 15 May 2020

Jones-Norse, Linda (PhD, Clinical Science)

Development of Crosswalks to Aggregate International Spinal Cord Injury Functional Data

Thesis directed by Gale Whiteneck and David Weitzenkamp

ABSTRACT

In spinal cord injury, there are multiple databases containing information on functional recovery, but data cannot be pooled or compared due to differences in how function is measured. A crosswalk (where scores from two separate assessments are linked or converted to a common metric) is needed to allow comparisons. Thus, the primary objective was to create a crosswalk between the Functional Independence Measure (FIM) and the Spinal Cord Independence Measure III (SCIM III) for items reflecting voluntary motor function.

Common person equating, in which the instruments are administered to the same group of individuals, was used to create and validate the crosswalks. Three databases were used: The Swiss Network of Spinal Cord Injury database (SWISS, n = 663) for development, the Rick Hansen Spinal Cord Injury Registry (RHSCIR, n = 557) and the United States based SCIM III reliability study (Anderson, n=390) for validation. Three conceptually different crosswalk methods were used. 1) Expert panel evaluation where experts establish equivalency for similar items and scores between measures, developing a third common scale to which individual FIM and SCIM III scores were then re-coded; 2) Equipercentile equivalency in which total scores on both measures are aligned by percentile rank, and 3) Rasch analysis where items from both scales are co-calibrated based on item difficulties.

Comparisons were made between the expert panel FIM and SCIM III crosswalk scores, and between the raw and crosswalked scores for equipercentile and Rasch crosswalks. All methods resulted in high correlation coefficients (0.897-0.971) across all databases. Additional analysis of score distributions,

distribution and amount of differences, point differences and sub-group invariance suggest that the equipercentile and Rasch crosswalks are the most accurate.

As the Rasch method creates a crosswalk using a linearized scale based on logits, which can be compared to other outcome measures, the Rasch crosswalk is suggested for use. This crosswalk can be used to allow comparisons of functional recovery across multiple databases reflecting different systems of care and rehabilitation approaches.

DEDICATION

I dedicate this work to both my immediate and extended family, who provided much needed support on my non-traditional PhD journey. People with SCI are the inspiration for all of the work we do and without their participation in research projects and registries to advance the field, this work would not be possible.

Gale Whiteneck provided such steadfast and consistent feedback, with continued patience. When Gale agree to work on the "FIM/SCIM III crosswalk" I am not sure he knew what he was in for! Dan Lammertse and Jan Tucker have been so very supportive of this journey, with Dan as a mentor in many ways in all stages of my career. Jan has been inspirational and seemed to know just when I needed a "pick me up" email or call. MJ Mulcahey always puts things in perspective, reminding me that my family is the priority and that I am so lucky to have the opportunity to pursue a PhD, even as I bemoaned writing another school paper on a Friday night! John Steeves has exposed me to so many new and interesting opportunities. Marion Murray served as a role model as a woman in science, who also knew how to have a great time!

My parents and step-parents provided the means and support for my college education. Although neither of my parents completed a university education, there was never a question that this was the path my brother and I would pursue. My parents supported out of state tuition for my first two degrees, which enabled me to graduate debt free, a true gift in this day and age. Without that, I doubt I would have pursued additional education.

I also wish to acknowledge the impact of fate. A skiing accident in my youth (Jackson Hole, WY) diverted my education from business (which I would not have enjoyed nor thrived in!) to physical

therapy. That path led to me circuitously to spinal cord injury research, where I have found a professional home.

And most of all to my husband Ed and my beautiful, creative and talented daughters. My husband has supported my entire career path from day 1 and picked up the pieces with the family, when I was not present. My daughters are my constant inspiration to be a better person. My nuclear family has sacrificed a great deal to support me and I dedicate this dissertation to them, first and foremost as well as my mom and aunt, who are not here to celebrate with me.

ACKNOWLEDGMENTS

This project was funded by the University of Texas Medical Branch, Center for Large Data Research and Data Sharing in Rehabilitation (sub-award # 18-84411-9). This project is supported in part by NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. Contents are the author's sole responsibility and do not necessarily represent official NIH views. The Rick Hansen Spinal Cord Injury Registry and this work are supported by funding from the Praxis Spinal Cord Institute, Health Canada, Western Economic Diversification Canada, and the Governments of Alberta, British Columbia, Manitoba, and Ontario. For more information about RHSCIR, please visit www.praxisinstitute.org.

Site Acknowledgements: Data for this research/project was collected using the Global Research Platform http://www.rhigrp.net/.

Neurology data for this research/project was cleaned using the RHI ISNCSCI Algorithm http://www.isncscialgorithm.com/.

Additional data sources included data from the US based SCIM III reliability study and a Swiss study comparing FIM and SCIM III in spinal cord injury, generously supplied by Kim Anderson and Armin Curt, respectively.

I also wish to thank John (Kip) Kramer, who identified the need for a FIM/SCIM III crosswalk, Chih-ying (Cynthia Li) who assisted with development and interpretation of the Rasch crosswalks, and Vanessa Noonan who participated in the expert panel method. In addition, physical and occupational therapists from Magee Rehab, Philadelphia, Pennsylvania, Rick Hansen Institute and GF strong in Vancouver, British Columbia, and Shepherd Center in Atlanta, Georgia generously donated their time to review and comment on the expert panel crosswalk.

Definition of Terms

ACT	American College Testing
ADL	activities of daily living
Anderson	validation dataset derived from the US based SCIM III reliability study[1]
ASIA	American Spinal Injury Association
B,F&S	body function and structure
CARE Tool	Continuity Assessment Record and Evaluation
CDE	common data elements
CFA	confirmatory factor analysis
CGI	clinical global impression
CMS	Center for Medicare & Medicaid Services
COA	clinical outcome assessment
CTT	classical test theory
EFA	exploratory factor analysis
EMSCI	European Multi-Center Study About Spinal Cord Injury
EFS FIM	expert panel FIM/SCIM III FIM
EFS SCIM	expert panel FIM/SCIM III SCIM
EQ FIM	equipercentile FIM
EQ SCIM	equipercentile SCIM
FA	factor analysis
FIM	Functional Independence Measure
ICF	International Classification of Functioning, Disability and Health
IRT	item response theory
ISNCSCI	International Standards for the Neurological Classification of Spinal Cord Injury
K-MBI	Korean Modified Barthel Index
LOS	length of stay
MCID	minimal clinically important difference
MDS	Minimum Data Set
MFIS	Modified Fatigue Impact Scale
MS	Multiple Sclerosis
NINDS	National Institute of Neurological Disorders and Stroke
NIDDILR	National Institute on Disability, Independent Living, and Rehabilitation Research
NIDRR	National Institute on Disability, and Rehabilitation Research
NSCISC	National Spinal Cord Injury Statistical Center
OT	occupational therapy
PRO	patient reported outcome

PROMIS	Patient-Reported Outcomes Measurement Information System
PT	physical therapy
R FIM	Rasch FIM
RHSCIR	validation dataset from the Rick Hansen Spinal Cord Injury Registry
ROC	receiver operator characteristics
R SCIM	Rasch SCIM
ROM	range of motion
SAT	Scholastic Achievement Test
SCI	spinal cord injury
SCI-FI	Spinal Cord Injury Functional Index
SCIM	Spinal Cord Independence Measure
SCIMS	Spinal Cord Injury Model Systems
SD	standard deviation
SMD	standard mean difference
SWISS	development dataset from four rehabilitation centers in Switzerland
US	United States

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER

I. INTRODUCTION

1.1 Overview

In different parts of the world, two 47-year-old men hit a patch of black ice and roll their cars: Steffen in Switzerland and Jack in the United States (US). Both of them incur a similar spinal cord injury (SCI) in terms of level and severity, and both receive surgical stabilization and decompression of their spinal column within 12 hours of injury. After they are medically stable, both are transferred to rehabilitation facilities. Both men wish to know how long they will remain in inpatient rehabilitation. Jack is told that the average length of stay for his level and severity of injury is 36 days, whereas Steffen is told his stay will be approximately 152 days. Due to the short length of stay, Jack's inpatient rehabilitation will focus on the skills he needs to be discharged safely to home and be as independent as possible, such as bowel and bladder management, and transfers from wheelchair to bed, toilet and shower chair (primarily compensatory techniques). If eligible and depending on his insurance and the availability of a SCI rehabilitation center near his home, Jack subsequently can work on motor training as an outpatient to promote recovery of function. As Steffan's inpatient rehabilitation stay will be significantly longer, he can focus not only on transfer training (compensatory) but also on motor training (recovery).

When Jack asks what his expected progress will be, compared to other parts of the world with longer lengths of stay, he is told that this information is not available as a different clinical outcome assessment (COA) for activities of daily living (ADLs) typically is used in systems of care for other countries. Both Jack and Steffen are eligible for a SCI research study. When they ask about the primary outcome for the study, they are told it is based on natural recovery patterns from a large European

database. Jack wonders if the primary outcome is applicable to him as he is aware the systems of rehabilitation care between Europe and the US are substantially different.

This example illustrates a gap in the field of SCI. In Europe, the COA used to assess ADLs is the Spinal Cord Independence Measure, version III (SCIM III), whereas in the US, a generic functional COA, the Functional Independence Measure (FIM[2]) is used. Thus, comparing and contrasting functional outcome measures in these two systems and between research studies using one or the other of these COAs is not possible. In addition, both primary endpoints and appropriate parameters for inclusion/exclusion criteria often are based on recovery derived from one data base or another, which may not be relevant in a different system of care. The development of a crosswalk between FIM and SCIM III will enable better comparisons and the establishment of relevant equivalencies.

1.2 Significance

As of 2013, the estimated global incidence of traumatic spinal cord injury ranged from 8.0 - 246.0 cases per million persons per year, with prevalence ranging from 236.0 – 1298.0 per million.[3] In the United States, the annual incidence of SCI in 2018 was 17,810, with an estimated prevalence of 294,000.[4]

These numbers, however, which are far less than other neurological disorders, do not adequately illustrate the significance of SCI to the individual or to society. The average age of an individual incurring a SCI in the US is 43, with an average life expectancy at this age of 30 years for people with paraplegia, and 22.5 – 26.5 years for those with tetraplegia.[4] SCI impacts motor and sensory systems such that mobility and activities of daily living often are impaired, with numerous secondary complications such as infections, cardiovascular dysfunction and dysregulation of bowel and bladder function. Depending on the neurological level of injury, a severe injury can result in ventilator and/or wheelchair dependence. SCI is a life-changing event impacting quality of life, community participation,

economic opportunity and results in significant life-long utilization of heath care services. At present, there are no drug or biological agents approved for neuroregeneration or neuroprotection leading to functional recovery following SCI. Therefore, the primary intervention to maximize physical function following SCI is rehabilitation. Functional improvements following rehabilitation are tracked using functional outcome measures. As it is important that a given clinical intervention improves how a person "feels, functions, or survives", functional outcome measures are a necessary clinical trial endpoint, particularly in pivotal (late stage) or Phase III clinical trials.

Ideally, universal therapeutic approaches can be identified to increase function, leading to improved health and well-being of individuals with SCI. Although many rehabilitation trials have been conducted, very few clinical trials for neurorecovery have been completed for SCI and the identified primary outcomes were not always directly clinically relevant.[5,6] Numerous clinical trials currently are underway, thus it is important to identify clinically relevant outcomes.[7] These outcomes must exceed the rate of spontaneous natural recovery following SCI, thus, it is necessary to quantify natural recovery using data from as many participants as possible, ideally across multiple databases, representing diverse systems of care.

In the field of SCI, there are multiple, independent databases containing information on neurological and functional recovery. The two largest databases, the United States based Spinal Cord Injury Model Systems (SCIMS) database and the European Multicenter Study about Spinal Cord Injury (EMSCI) database, use similar but slightly different COAs (FIM and SCIM III, respectively), thus directly aggregating or comparing these data is not possible. In addition, published SCI studies typically use either FIM or SCIM, thus one cannot compare functional outcomes across studies.

The solution is to create a "crosswalk" or link between FIM and SCIM III scores to enable comparisons and establish equivalencies between databases and research studies.

1.3 Background

Development of crosswalks or score linking originally was used in the field of education to establish equivalent scores in standardized tests. An oft-cited example in education is that of linking scholastic aptitude test (SAT) scores with those of the American College Testing (ACT), so that regardless of which test a high school student takes, scores can be compared. In recent years, these methods also have been applied to COAs.

Linking is appropriate when tests assess a similar construct but with different test construction (e.g. number of items and scoring). Examples in health care outcomes are that of a short vs. long form of an assessment such as the Spinal Cord Injury Functional Index (SCI-FI), a patient reported functional outcome measure, or measures that are intended to assess change across the lifespan, such as linking the SCI-FI and the Pedi SCI, a pediatric patient reported functional outcome measure. In this case separate but related measures may be used depending on developmental stage. Test construction differs with fewer questions and different scaling, but the underlying construct is intended to be similar. Other cases where linking is appropriate is when a newer measure is developed, assessing a similar construct to a prior measure. In all of these situations, a crosswalk table can be created, where scores on one test or measure are “linked” to scores on a second measure.

Three conceptually different methods can be used for linking, with different strengths and assumptions. Expert panel linking is a method whereby experts in the field establish relative equivalency for similar items and scores across instruments based on their expertise and experience. Equipercentile linking is a commonly used alignment method, in which a crosswalk is developed based on aligning total score distributions and rank ordering total scores for both measures. Linking methods using Rasch analysis co-calibrates items on a common linear scale and creates a crosswalk based on item difficulty,

related to the underlying construct measured by the COAs using more modern and sophisticated statistical approaches. Additional details for these methods are found in Chapter 2 (Section 2.9).

After a crosswalk is created, the strength of the crosswalk should be assessed. Dorans[8] and Dorans and Holland[9] provide three fundamental criteria that assess the degree to which measures can be linked in a crosswalk: (1) construct similarity, (2) the strength of the observed (empirical) relationship between the scores to be linked, and (3) population invariance. Although these are identified as separate criteria, they are tied to construct similarity.

Construct similarity between COAs is a core requirement for a successful crosswalk. This can be assessed by comparing COA content, and/or statistically using factor and/or Rasch analysis (described in Chapter 2). If multiple constructs are noted within COAs, it may be necessary to create crosswalks for each construct. The decision to create multiple or single crosswalks should also reflect how the crosswalk will be used.

The empirical relationship between two measures primarily is assessed via the correlation between actual and cross-walked scores. Additional assessments of the relationship between measures include similar score distributions and point differences between actual and cross-walked scores. Although a strong relationship, as assessed by correlation, does not necessarily reflect construct similarity (therefore review and comparison of the items, descriptors and scoring is also important), a weak correlation implies construct differences.

Population invariance considers where differences in populations (for example male and female) between the actual and cross-walked scores may exist. If the measures assess a similar construct, population invariance will be similar between actual and cross-walked scores. For example, if a test is intended to assess stride length, the differences between males and females will be similar between

actual and cross-walked scores, e.g. as females are shorter, female stride length will be shorter in both the actual and cross-walked scores. If this is not the case, the test is likely evaluating something other than stride length, thus the underlying construct between the measures is different.

Dorans and Holland[9] consider the additional criteria of reliability and symmetry to be less critical in assessing crosswalk strength. They note that measures have been linked with dissimilar reliability, in which a strong crosswalk is still created. However, if unequal or low reliability is present and if correlations are not strong between actual and cross-walked scores, it is difficult to discern if the weak correlation is due to construct differences or error due to low reliability. Although the criteria of a symmetrical linking function (A to B is the same as B to A) appears intuitive, in some linking methods (such as equipercentile linking) this criterion will not precisely hold due to rounding.

FIM and SCIM III appear to assess a similar construct, ADLs (albeit in a slightly different way), and have similar reliability, providing support for this crosswalk approach.

1.4 Research Objectives and Rationale

The **primary objective** of the proposed work was to create a crosswalk (where scores from two assessments are linked or converted to a common metric) across FIM and SCIM III items reflecting voluntary functional movement. For this project, items reflecting functional movement were defined as items associated with voluntary movement contributing to independence in ADLs. The **rationale** for the proposed work is that a crosswalk is needed to enable more direct comparisons between all SCI databases. To address this gap we used retrospective data from three databases, the Swiss SCI Centers (SWISS), the Rick Hansen Spinal Cord Injury Registry (RHSCIR) based in Canada, and the US- based SCIM III reliability study, in which both FIM and SCIM III data were collected on a sub-set of patients. The largest of these (SWISS, n =663 participants) was used as a linking dataset to establish crosswalks using three conceptually different methods, while the smaller datasets (RHSCIR, n = 558 and US- based SCIM

III, n= 390) were used to validate these methods. Although these datasets are from different countries, with different rehabilitation approaches and lengths of stay, the data are used to compare FIM and SCIM III at a given time point, thus the differences between datasets are not relevant. The output of this project **significantly** contributes to the field of SCI rehabilitation, as a valid crosswalk is the fundamental first step toward leveraging comparable information from multiple, large databases to better understand recovery of function following SCI.

1.5 Specific Aims and Hypothesis

Specific Aim 1: Assess the number of dimensions* in a combined FIM and SCIM III voluntary motor function item bank.

Hypothesis: Assessments of dimensionality will support the use of a single crosswalk for each crosswalk method.

Null hypothesis: Assessments of dimensionality will support the use of multiple crosswalks for each method.

*Dimensions refers to the number of underlying constructs, traits or domains (used interchangeably), defined for this purpose as a behavioral or physiological characteristic.

Specific Aim 2: Crosswalk(s) for FIM and SCIM III voluntary motor function items will be created using three conceptually different methods: expert panel linking, equipercentile linking and Rasch analysis co-calibration. Correlations between actual and cross-walked scores using the crosswalk(s) for each of the three methods will be assessed.

Hypothesis: Correlations will exceed established criteria (0.866)[8] using the crosswalk(s) for each of the three methods.

Null Hypothesis: Correlations will not exceed established criteria (0.866)[8] using the crosswalk(s) for each of the three methods.

Specific Aim 3: Validate the crosswalk(s) for each of the three methods in separate datasets.

Hypothesis: Correlations in the validation dataset will exceed established criteria (0.866) using crosswalk(s) for each of the three methods.

Null Hypothesis: Correlations in the validation dataset will not exceed established criteria (0.866) using crosswalk(s), for each of the three methods.

CHAPTER

II. REVIEW OF LITERATURE

2.1 Introduction

In this chapter, a summary of relevant literature is presented. First, a general background in outcome assessment in SCI with specific descriptors and background information on FIM and SCIM III, and a comparison of the two measures is provided. A summary of major SCI databases and how information about functional recovery from these databases has been used, highlighting a key gap and therefore the importance of developing a crosswalk between FIM and SCIM III is presented. Methods to develop crosswalks and their theoretical foundation in relation to classical vs. modern psychometric theory are discussed. Examples and analyses of how these methods are used in health care research are provided. Methods to assess crosswalk strength and validation also are discussed.

2.2 Outcome Assessments in SCI

As in research for any type of illness or disability, there are numerous COAs for clinical and research use in SCI. Following the terminology of the International Classification of Functioning, Disability and Health (ICF), these range from assessments of body functions and structure (BF&S), to activity and participation.[10] As summarized by Jones et al.[11], in neurological disorders, assessments of B,F&S measure impairment based on neurological and physiological outcomes measures, are typically used in the early phases of a research investigation. For example, in Phase I (first in human, safety, tolerability and feasibility) and Phase II studies (safety, efficacy, proof of concept) measures of impairment such as muscle strength and sensation, and electrophysiology and/or reflexes might be used to assess safety, and/or evidence of biological activity. However, measures of impairment do not indicate functional benefit. In late phase studies, where it is important that a given therapeutic improves

how a person “feels, functions, or survives”, functional outcome measures capture this information. These measures can focus on a specific intervention such as locomotor training, where one might assess walking speed or distance, or performance of ADLs such as FIM or SCIM III.

The National Institute of Neurological Disorders and Stroke (NINDS) has developed a list of common data elements, based on expert review, of COAs that are recommended for use in SCI research.[12] The goal is to enable use of standard data elements for comparisons across research studies. In addition to categories such as demographics, epidemiology, and hospital care/management, 117 “outcomes and endpoints” are identified as well as 28 International SCI Data Sets. Of the outcomes and endpoints, 41 functional outcome measures are listed, which encompass impairment and function (specific and general such as ADLs). All CDEs are ranked based on the strength of published evidence and applicability for research.

Although FIM is not listed in the NINDS CDEs, it is one of the two most commonly used assessments of ADLs in SCI (Section 2.2.1 below), along with the SCIM III. A summary of FIM and SCIM III including instrument description, development and psychometrics is provided below.

2.2.1 Functional Independence Measure

The Functional Independence Measure (FIM) was developed in 1984 to assess burden of care, document disability and measure functional outcomes following rehabilitation, in a consistent way.[13] FIM is the most widely used generic functional outcome assessment tool and currently is collected as part of the Inpatient Rehabilitation Facilities Quality Reporting Program in the United States. As this measure is proprietary, a copy is not provided. A summary/overview of the FIM (and SCIM III) can be found in Table 1.

The FIM is an ordinal scale consisting of 18 items in motor (13 items) and cognitive domains (5 items). Scoring is "Likert-like" from 1 – 7 where a score of 1 reflects "total assistance" and a score of 7 represents "independence". Scores below 6 require assistance or supervision. FIM may be administered by observation (30-45 minutes) by a multi-disciplinary team of physical therapy (PT), occupational therapy (OT) and nursing, by questionnaire or self-report. Normative data based on SCI level and completeness of injury were established for a variety of time points following injury. [2] Use of FIM requires a license and there is an associated fee. It is available in 11 languages as well as in a pediatric version.[14]

As noted, FIM is an ordinal measure. Ordinal measures rank scores from lowest to highest or easiest to hardest, but differences between scores can vary and are not measureable. One can determine that someone has more or less of a given trait but cannot quantify the difference mathematically. As such, ordinal data are described using median (middle value in a list of numbers) and mode (the value that occurs most often) which considers the order of data, versus the mean (average) which is used for interval level data where the distance between two scores is known and is the same. Although a mean can be calculated for ordinal data, it is not recommended and in some cases is meaningless as the distance between scores differs, so the average is not relevant. Statistical methods consider the type and distribution of data. For ordinal or non-normally distributed data, non-parametric statistics are recommended as they do not use mean and standard deviation as the underpinning of the analysis, but use the position of scores. Parametric statistics are used for interval level data that are normally distributed or with a sample size greater than 30. Although parametric statistics are often applied to ordinal data, the appropriateness and interpretation must be undertaken with caution.

Table 1: Summary of FIM and SCIM III items

FIM (7 levels for all items)	SCIM III	# of levels for SCIM III items
Voluntary Motor Items		
Eating	Feeding	4 (0,1,2,3)
Grooming	Grooming	4 (0,1,2,3)
Bathing	Bathing upper body	4 (0,1,2,3)
	Bathing lower body	4 (0,1,2,3)
Dressing - upper body	Dressing upper body	5 (0,1,2,3,4)
Dressing - lower body	Dressing lower body	5 (0,1,2,3,4)
Toileting	Use of toilet	5 (0,1,2,4,5)
	Mobility in bed	4 (0,2,4,6)
Transfers: bed,chair,wheelchair	Transfers: bed - wheelchair	3 (0,1,2)
Transfers : toilet	Transfers: wheelchair- toilet-tub	3 (0,1,2)
Transfers: tub,shower		
Locomotion: walk/wheelchair	Mobility: Indoors	9 (0,1,2,3,4,5,6,7,8)
	Mobility: Moderate distances	9 (0,1,2,3,4,5,6,7,8)
	Mobility: Outdoors	9 (0,1,2,3,4,5,6,7,8)
Locomotion: Stairs	Stair Management	4 (0,1,2,3)
	Transfers: wheelchair-car	3 (0,1,2)
	Transfers: ground - wheelchair	2 (0,1)
Autonomic items		
	Respiration	6 (0,2,4,6,8,10)
Bladder management	Sphincter management-bladder	7 (0,3,6,9,11,13,15)
Bowel management	Sphincter management --bowel	4 (0,5,8,10)
Cognitive Items		
Comprehension		
Expression		
Social interaction		
Problem solving		
Memory		

Reliability and validity of FIM have been assessed in mixed and disease specific populations. As this project is in SCI, the focus was on SCI specific literature or mixed diseases in which individuals with SCI were included, if SCI specific literature does not exist. Additionally, the motor and cognitive FIM sub-

scales assess different domains and the cognitive sub-scale often is not relevant in SCI. Where possible, we will report motor FIM analyses from the literature or total scores where necessary.

Face validity is a basic criterion that assesses whether an instrument "looks like" it measures what it purports to measure. Content validity refers to whether an instrument covers all facets of the construct it is intended to measure. Both criteria are somewhat subjective and typically are addressed in stages through the test development process, based on expert input. Face and content validity for FIM were incorporated during the FIM development process[15], and in subsequent evaluations.[16]

Construct validity assesses whether an instrument measures the underlying construct it is intended to measure. This can be assessed statistically by evaluating items within the instrument and comparing to other measures, assessing a similar construct. Two methods for assessing items within a COA are: 1) factor analysis, which explores the degree to which individual tests items are related to underlying constructs that are not directly measured by the instrument, and 2) Rasch analysis which considers individual item fit to a single underlying construct or dimension. More details on these methods are provided later in this chapter (Section 2.9.3 and 2.9.6). In 1993 Heinemann et al.[17] and Stineman et al.[18] clearly demonstrated that FIM contains a cognitive and motor construct in mixed population studies, using Rasch and factor analysis. Stineman et al.[18,19] further assessed FIM's underlying construct(s) in a mixed population by looking at impairment-specific dimensions using factor analysis. In the SCI impairment group, three underlying factors were identified in the complete FIM assessment. Based on the activities that correlated with those factors, the underlying constructs appeared to be cognitive factors and ADL factors related to use of the upper extremities or lower extremities, reflecting deficits primarily associated with tetraplegia and/or paraplegia. In a recent publication by Hong et al.[20], in a general outpatient rehabilitation population, items from the FIM and the Korean version of the

Modified Barthel Index were assessed in a single item bank using factor and Rasch analysis. Three constructs were identified: self-care, mobility and involuntary movement.

Construct validity also is assessed by the degree to which an instrument is related to other instruments with similar constructs. As summarized in rehabmeasures.org[21], FIM has been correlated with numerous measures in individuals with SCI ranging from impairment based measures, to walking, upper extremity function and client centered assessments.[22-28] Study designs varied, where FIM total motor, locomotion or upper extremity collected by observation or questionnaire were correlated with other instruments. Correlations range from weak (client centered) to very strong (upper extremity function, motor scores), with most correlations in the moderate to strong range, when using the statistical criteria of Schober et al.[29](0.00–0.10 = negligible correlation, 0.10–0.39 = weak correlation, 0.40–0.69 = moderate correlation, 0.70–0.89 = strong correlation, 0.90–1.00 = very strong correlation).

Responsiveness of an instrument reflects the sensitivity to detect change over time. In the SCI population, FIM was found to be responsive to change [1,30-32] with reduced responsiveness for some of the mobility items.[33]

The *minimally clinically important difference* (MCID) reflects the smallest amount of change that is clinically meaningful to an individual. MCID of FIM for individuals with SCI has not been assessed. In stroke, the MCID for motor FIM was found to be 17 points. *Normative data* for FIM has been established for individuals with SCI.[2]

Floor and ceiling effects reflect the degree to which an instrument assesses lower and higher levels of function. For example, in an individual with a more severe injury, such as a complete, high level cervical injury, are there items which reflect this lower level of function? Floor and ceiling effects related

to severity and level of injury were identified for FIM, meaning the instrument does not have sufficiently easy and hard items to capture function at less and more severe injuries.[2]

Numerous FIM *reliability* studies have been conducted. In 1996, Ottenbacher et al.[34] summarized the reliability studies completed to that point. Eleven studies were identified, with mixed and disease-specific populations. Studies focused on interrater reliability and/or equivalence reliability (e.g. equivalence between two modes of administration), with correlations calculated by intra-class correlation or Kappa coefficients. FIM was collected by a multi-disciplinary team or a single discipline. Correlations for total FIM scores ranged from 0.83 – 0.99. Median reliability values for individual motor items ranged from 0.66 (stairs) to 0.90 (toilet transfer), with an overall average correlation of 0.97. Authors concluded the FIM has "good interrater reliability across a wide variety of raters with different professional backgrounds and levels of training".

A single study reviewed by Ottenbacher et al. [34] was specific to SCI. Interrater reliability of observed FIM's was calculated, with FIM collected by a multi-disciplinary team of PT, OT and nursing.[35] Segal et al.[35] assessed interrater and inter-institutional reliability in a small study of 57 participants transferring from acute care to rehabilitation, with a maximum of six days between observations. Reliability as assessed by Pearson correlation coefficient was 0.83 for total scores and ranged from 0.10 – 0.72 for individual motor FIM items. The proportion of agreement for individual motor FIM items ranged from 0.32-0.95. Reliability based on impairment category (level and extent of injury) also were calculated: complete tetraplegia (n=14, 0.870), complete paraplegia (n=13, 0.74), incomplete tetraplegia (n=17, 0.49), and incomplete paraplegia (n=9, 0.85). In the total sample, items with above average reliability for both Pearson and proportion agreement were feeding, bed, toilet and tub transfers. Items below average for reliability and agreement were part of the cognitive sub-scale. It is notable that the lowest correlation was in incomplete tetraplegia, which is the most common category of SCI (47.2%).[36]

This may be due to the wide range of impairments (upper and lower extremity, trunk) impacting ADLs within this category.

One study compared FIM clinician and self-care ratings, while another compared clinician ratings by observation and questionnaire. Grey and Kennedy[37] assessed 40 participants and demonstrated a high correlation (type not noted) for total scores of 0.828 between the two modes of administration, with no significant differences for either total or sub-scale scores. For the motor sub-scales of self-care, sphincter control, mobility and locomotion, the correlations ranged from 0.454 (locomotion) to 0.841 (self-care). FIM by observation was obtained by nursing staff every six weeks from admission to discharge. A FIM questionnaire was sent to study participants at frequent intervals and the final inpatient FIM and the first post-discharge FIM then were compared, with an average time between ratings of 7.25 weeks. Authors concluded that FIM self-report was reliable. Karamehmetoğlu et al.[38] compared FIM by observation vs. clinician questionnaire in 50 participants with a total score Spearman correlation coefficient of 0.94. Not all individual item correlations were presented. Of those discussed the range for correlation of motor items was 0.65 (upper body dressing) to 0.97 (bowel management). A single clinician completed FIM by questionnaire and then by observation. Both of the above studies were small and in the second study a single examiner completed both methods, thus the findings are prone to bias. Reliability of the alternative method (questionnaire or self-report) was not conducted, thus it is not possible to determine if self-report and clinician questionnaire are viable methods for individuals with SCI, nor draw conclusions about the "best method".

Internal consistency, as measured by Cronbach's alpha, is considered to be a measure of reliability, specifically of the relatedness of individual test items. A high correlation indicates individual test items are correlated with each other; however, this interpretation is influenced by the fact that correlation will increase with the number of test items. Internal consistency is linked to assessment of

unidimensionality, or construct, but is only part of the assessment, where factor analysis and Rasch analysis (Section 2.9.4, 2.9.4.1) are more direct measures of dimensionality. Internal consistency is often assessed during test development to determine relatedness and redundancy of test items. It is related to reliability in that it assesses reliability of test items with the test itself and validity as it relates to assessment of underlying construct. For clinical applications, a minimum value of 0.90 is suggested.[39] Internal consistency for motor FIM has been reported to range from 0.934 – 0.953 in SCI.[18,40]

Despite small studies with variable findings, and some sub-optimal study designs, FIM is considered to be a valid instrument. In the single study assessing interrater reliability in SCI, reliability was high for total scores, ranging from weak to strong for individual items, with low to high percent agreement for individual items. However, as total scores typically are used, the influence of individual items is lessened. FIM is a widely used and accepted measure of ADLs.

FIM data collection was required by the Centers for Medicare & Medicaid Services (CMS) for all inpatient rehabilitation units or hospitals as an assessment of standard of care. In 2005, the Deficit Reduction Act directed CMS to track health status across the continuum of care, from inpatient hospitalization to rehabilitation and long term or home health care. As a result, the Continuity Assessment Record and Evaluation (CARE) Item Set was developed and was slated by CMS to replace FIM on 1 October 2019. Although the ADLs assessed in both COAs are similar, variables have been added, some descriptors have changed and scoring is now 1-6 (vs.1-7).[41] Despite the transition to the CARE assessment, establishing a FIM/SCIM III crosswalk is valuable for comparing historical data such as that of the SCIMS, which began collecting motor FIM data in 1988. In addition, creation of a FIM/SCIM III crosswalk using three diverse methods may be useful in constructing a SCIM III/CARE tool crosswalk. Additionally, US researchers may continue using FIM until the psychometrics of the CARE tool are

established. To date, no FIM/CARE crosswalk for SCI has been created, although such a crosswalk for traumatic brain injury is underway.

2.2.2 Spinal Cord Independence Measure

SCIM was developed as a SCI specific COA with the initial publication in 1997.[42] Developers note that SCIM is intended to measure the effectiveness of rehabilitation training and is specific to SCI function. Since the development of the original SCIM, it has undergone two revisions, resulting in SCIM II and III with a fourth version currently in the validation process.[32,43] The goals in developing the SCIM were to: 1) include only items relevant to function following SCI; 2) incorporate weighting of items and scoring to reflect the importance of the activity; and 3) precisely define scoring criteria. A summary of SCIM III items is in Table 1. The complete SCIM III version can be found in Appendix A.

The original SCIM consisted of three sub scores: self-care, respiration/sphincter management, and mobility with a total of 16 items.[42] Mobility was further divided into room, toilet, indoors, moderate distances, and outdoors on even surfaces. The maximum score was 100 with a minimum of 0. The current version of SCIM contains 17 items, with the same sub-scales and maximum possible score; however item scoring is weighted and differs per category, with some items "skipping" scores.[32] For example, "mobility in bed" has possible scores of 0, 2, 4 and 6, while mobility indoors scores range consecutively from 0-8. SCIM III can be administered by observation (takes 30- 45 minutes), interview or self-report, with the highest reliability by observation, which typically is conducted by a multi-disciplinary team of PT, OT and nursing.[44] SCIM III is non-proprietary, freely available and has been translated into three languages. A self-report version for youths with SCI is also available.[45]

Significant changes during the SCIM development process included separating upper and lower body functions in the bathing and dressing category, as this better reflects a major difference in the functional abilities of individuals with SCI (paraplegia vs. tetraplegia). Descriptors for given items and

scoring also has been revised and a ground to wheelchair transfer item was added to SCIM III. Each version of the SCIM has undergone psychometric evaluation, the findings of which have informed future versions.

As with FIM, *face and content validity* have been considered throughout the development process, based on expert input.[32,42,43] *Construct validity* in terms of instrument dimensionality, was assessed by Catz et al.[46] by sub-scale (vs. individual items), using factor and Rasch analysis. The analysis was performed on sub-scales vs. individual item. Factor analysis supported a unidimensional scale, while Rasch fit statistics indicated some items with misfit which may indicate multi-dimensionality. More details on the Rasch analysis are provided in the section (2.9.5) below.

Construct validity has been assessed through comparisons with numerous instruments addressing impairment (upper extremity motor function), walking function and speed, balance, upper extremity capacity tests and subjective independence.[26,47-50] Correlations were weakest in relation to subjective scores of independence (as assessed by individuals with SCI), and generally strong in measures of impairment and capacity. SCIM III is responsive to change based on several studies.[1,32,51]

SCIM III was found to be *responsive* to change in acute rehabilitation[1,32] and a sub-acute day program.[51] *Normative data* are available for SCIM III.[52] *MCID* and smallest real difference (the smallest change reflects true change vs. measurement error) were assessed in a study by Scivoletto et al.[53] A distribution based approach was used, based on retrospective chart review. For the total SCIM score, a change of four points was identified as a small significant improvement, while a change of 10 points was identified as substantial improvement. The smallest real difference, however, was approximately 8-10 points across databases, meaning that a clinically significant change of 10 points is just at the point where true change is detectable. MCID also was broken down by sub-scale, in which the respiratory and sphincter management sub-scale had the largest contribution to MCID. MCID was not broken down by

voluntary motor scores, which would include the self-care, mobility room and toilet, mobility indoors and outdoors on even surface sub-scales, and the toilet transfer item within the respiration and sphincter management sub-scale.

SCIM III *internal consistency* as assessed by Cronbach's alpha, ranges from 0.770- 0.849.[32] *Floor and ceiling effects* have been observed.[46,51]

Reliability studies of the English version of SCIM I, II, and III are summarized in Table 2.[1,32,42,43] In all studies, similar methods were used. SCIM III data for individuals with traumatic and non-traumatic SCI were collected within the first week of admission and within one week prior to discharge from rehabilitation. SCIM III data were collected by observation by two staff members with relevant expertise. If these data were not directly observed (e.g. specific self-care or sphincter tasks), information was collected from relevant staff members who had observed these activities. The first two versions of SCIM[42,54] were assessed in small (n=32 and 28) single center studies, run by the instrument developer, while later studies of SCIM III[1,32] were larger (n=425 and 463), multi-center studies in different countries and systems of care. Interestingly, the second validation study for SCIM III was based on recommendations from an expert panel at the National Institute on Disability and Rehabilitation Research (NIDRR, now National Institute on Disability, Independent Living, and Rehabilitation Research - NIDILRR) [55], who suggested a second study on reliability and validity of SCIM III should be conducted in the United States (US).[1] The rationale was that the US health care system differs substantially from other countries in which the prior study occurred and that to date, all assessments of SCIM had been conducted by the SCIM developers. In addition to assessment of reliability and validity, an additional aim of the study was to determine if instructions for assessors were necessary.

Table 2: Summary of SCIM observational psychometric studies

SCIM Version	Sample size and number of study centers	Percent agreement total score and individual tasks	Kappa coefficient for individual tasks	Pearson correlation for total and sub-scales scores	Intra-class correlation coefficients for total and sub-scale scores	Internal consistency (Cronbach's alpha)
SCIM I[42]	32, single center	72-99%, 12/16 ≥ 80%	0.66 - 0.98	0.98, 0.91-0.99	NA	NA
SCIM II[43]	28, single center	64-100%, 13/18 ≥ 80%	0.44 - 0.95	0.99, 0.90-0.97	NA	NA
SCIM III[32] **(Israel and Europe)**	425, 13 centers	74.5-96.2%, 13/18 ≥80%	0.63 - 0.82	0.94, 0.90-0.96	0.98, 0.95-0.97	All subscales > 0.70 (respiratory was 0.71)
SCIM III[1] **(US)**	463, 19 centers	65-99% (admit), 8/19 ≥ 80%, 67-94% (d/c), 11/19 ≥ 80%	0.56- 0.85 (admit) 0.60-0.81 (d/c)	0.91(admit)/ 0.96 (d/c), 0.81 (admit) /0.89 (d/c)	0.91 (admit) 0.95 (d/c) > 0.81 for all	All subscales > 0.70 except respiratory (0.61 at admit, 0.73-74 at d/c)

*d/c= discharge

Although study site training occurred for the SCIM III reliability studies, no formalized training exists, nor is there an accompanying manual. The developers believe that the descriptors are self-explanatory and no additional clarification is needed. Authors of the US based SCIM III study proposed that the higher degree of variability between raters was due to the lack of SCIM III scoring instructions on the worksheet, with the largest area of questions from researchers being in the description of "partial assistance". They suggested splitting this category into percentages such as "50% or more assistance vs. 50% or less assistance". In the current scoring metric, an individual requiring supervision would have the same score as an individual requiring maximal assistance. The proposed change would equate to categories 1 and 2 and 5 and 6 in the FIM. This would also allow for greater gradation when assessing minimal clinically important difference (MCID). A post- acute day program study also noted the need for a SCIM III manual.[51]

SCIM III also has been assessed for collection by interview and compared with observational data.[44] Although assessment by observation, which assesses the capacity of an individual to complete a task in a structured environment, may be more objective than by interview or questionnaire, it may not reflect performance which is what individuals do in daily life in their home and community. [11] In addition, assessment by observation may not always be feasible, particularly in longitudinal studies. To explore this, 35 individuals from Loewenstein Rehabilitation Hospital were administered SCIM III by interview and by observation. Researchers conducted inter-rater reliability of SCIM III obtained by interview and compared interview versus observation. Total agreement between interviewers ranged from 32-100%, with Cohen's kappa coefficient ranging from 0.11 – 0.80. Areas with the lowest correlation were in dressing, mobility in bed and mobility outdoors. Authors noted that the score differences varied by only one point in almost half the tasks. The intra-class correlation coefficient (ICC) was 0.88 for total SCIM score. Percent agreement between SCIM III collected by observation vs.

interview ranged from 17-100% with Cohen's kappa of 0.1-0.7. Again, the vast majority of differences were a difference of one point (90-100%). ICC for total SCIM was 0.905. Authors concluded that assessment by interview should be approached with caution. Although total score reliability was strong to very strong, as SCIM III frequently is collected by interview in registries and clinical research, the low individual item correlations are concerning.

A self-report version of SCIM III in German for inpatients was found to have a stronger relationship with SCIM III collected by observation.[56] Pearson correlations for total score were 0.87, with sub-scales ranging from 0.81 (respiratory and sphincter management) to 0.87 (all other sub-scales). ICC's also were strong with 0.90 for total scores and 0.83-0.86 for sub-scale scores. It was noted that self-report scores typically were higher than observed scores.

In summary, SCIM III is considered a reliable and valid measure of ADL function in SCI. Although global assessments of interrater reliability have not changed significantly over the SCIM iterations, the content changes have improved the clinical utility of the measure. Currently, it is the most commonly used COA assessing ADLs in SCI research studies, as reflected in the Spinal Cord Outcomes Partnership Endeavor SCI clinical trials tables (Rehabilitation and Technological Interventions to Improve Functional Outcomes, and Drug, Cell, and Surgical Interventions to Improve Neurological and Related Functional Outcomes).[57]

2.2.3 FIM and SCIM III comparison

Although both FIM and SCIM III assess activities of daily living, the intent, as described by the instrument developers of these COAs, varies. FIM is a generic COA to assess burden of care, while SCIM III was developed specifically for SCI and developers state it is intended to focus on the effectiveness of rehabilitation. There are a number of differences in items and scoring between the FIM and SCIM III. The complete version of FIM includes a motor and cognitive sub-scale, while SCIM III only includes motor

function. In most SCI studies, the cognition sub-scale is not used as this is not a primary deficit in an isolated SCI. Differences between the motor FIM and the SCIM III also are found in items capturing autonomic nervous system function (respiration and sphincter management). FIM does not include respiratory function, which can be impacted significantly following cervical SCI, at times necessitating ventilator support. Bowel and bladder dysfunction is a significant concern in individuals with SCI, with a profound impact on quality of life. Both measures contain items on bowel and bladder function, however these sections in SCIM III were developed with the specific needs and priorities of individuals with SCI in mind.

As noted, both FIM and SCIM III can be reduced to items reflecting voluntary motor function. Item differences in voluntary motor items are summarized in Table 3.

FIM scoring is consistent, with a scale of 1-7 across all items, where SCIM III scoring differs per item and is differentially weighted. For the novice assessor, remembering scoring criteria for SCIM III may be challenging. The advantages of FIM are that scoring is consistent and simple, and extensive training and certification on FIM is required every two years. Supportive manuals with additional clarification also are available. SCIM III is used increasingly in SCI studies in the US and elsewhere. As SCIM III rarely is used outside of the research setting in the US, even with study specific training, assessors are not as proficient as those who use SCIM III as part of their daily practice. As noted in studies cited above [1,51] and experienced by this author, the lack of a standardized manual for SCIM III is problematic. Industry and investigator sponsors of multi-center studies have created study specific manuals. The downside is that each manual is different, which may impact outcomes.

Table 3: Differences in FIM and SCIM III voluntary motor function items

FIM/SCIM III item	FIM	SCIM III
Eating/Feeding	No impactful differences.	
Grooming	No impactful differences.	
Bathing	Upper and lower body bathing are combined. Scoring is divided into 10 body parts, with each accounting for 10% of scoring.	Upper and lower body bathing are separate items.
Dressing upper body		Differentiates clothes with buttons, zippers and snaps in scoring.
Dressing lower body		
Toileting/Use of toilet	No impactful differences.	
Mobility in bed	No equivalent item in FIM.	
Transfers: bed, chair, wheelchair/bed-wheelchair	No impactful differences.	
Transfers: Toilet, tub, shower/wheelchair, toilet, tub	Toilet, tub and shower transfers are separate items.	Wheelchair, toilet and tub transfers are combined.
Locomotion (walk, wheelchair)/Mobility	Separates mobility items into walk and wheelchair. Distances for locomotion are specified by score and vary from SCIM III (<15 meters, ≥ 15 meters, ≥ 50 meters, ≥150 meters).	Three mobility items (indoors, moderate distances and outdoors) all of which encompass wheelchair vs. walk and include walking aids. Distances are specified per mobility item and vary from FIM. (10-100 meters, >100 meters)
Stairs/Stair management	Distances differ from SCIM III: < 4 stairs, 4-6 stairs, one flight of stairs, ≥ one flight of stairs.	Distance for all scores is ≥ 3 stairs.
Transfers: Wheelchair to car	No equivalent item in FIM.	
Transfers: Ground to wheelchair	No equivalent item in FIM.	

As noted, NINDS has developed a list of common data elements, based on expert review, of COAs that are recommended for use in SCI research to enable use of standard data elements for comparisons across research studies.[12] COAs are ranked based on the strength of published evidence and applicability for research. Core measures are those that are essential in any SCI study. Supplemental – highly recommended measures are strongly recommended, supplemental measures that commonly are collected, depending on study design, while exploratory measures require additional validation. While SCIM III is considered supplemental-highly recommended, FIM is not in the list of measures to be considered for use in SCI research, due to its lack of sensitivity to this specific condition (personal communication, Kim Anderson, January 2019).[1,32] The FIM for Children (Wee FIM) is rated as supplemental while the SCIM III Self Report for Youth is considered exploratory.

In comparing psychometric studies of both instruments, content, face and construct validity have been reported and considered acceptable. For construct validity, different instruments were correlated to FIM and SCIM III so similarities cannot be assessed directly. Although caution must be used in comparing reliability between FIM and SCIM III studies as study designs differed, reliability appears to be higher for SCIM III when assessing individual tasks as does percent agreement among raters on individual tasks. Total score reliability is strong for both measures, 0.83 for FIM (includes cognitive items) and 0.91-0.96 for SCIM III. SCIM III reliability studies used larger sample sizes (n=57 for FIM[35], n=425[32], 463[1] for SCIM III), which may have resulted in stronger reliability.

In two US SCIM III reliability studies[1,32], authors also assessed responsiveness and found that while FIM and SCIM III share "common components in the underlying construct", SCIM is significantly more responsive to change. In both studies, SCIM III was more responsive in sphincter management and respiration which were created specifically for SCI, and in the Itzkovich[32] study mobility indoors and

outdoors was also more responsive in SCIM III than FIM, although it is unclear to which FIM locomotion mode (walk, wheelchair or both) the comparison was made.

Dimensionality of both instruments using factor and Rasch analysis, shows mixed results. While FIM clearly contains a cognitive and motor component[17,18], studies have noted additional dimensions within the motor sub-scale.[19,20] A single SCIM III study[46] suggested a single underlying construct (despite individual item misfit), but the study design (evaluation of sub-scales vs. individual items) may have impacted these findings.

Correlations between FIM and SCIM III were calculated in the two SCIM III reliability studies, with values ranging from 0.779 – 0.80.[1,32] This moderate correlation conceptually supports the creation of a FIM/SCIM III crosswalk and suggests the COAs are measuring a similar construct. Correlations are likely to be higher when involuntary motor items are removed (items reflecting autonomic function) from both measures (respiratory, bowel and bladder sphincter management).

In summary, based on the similarity of constructs, moderate correlations between instrument total scores, and moderate to high instrument reliability, development of a crosswalk is supported.

2.3 Spinal Cord Injury Databases

Two of the largest and most commonly used SCI databases, SCIMS and the EMSCI, use the FIM and SCIM III respectively. At present one cannot compare functional outcomes between these two major databases, in the absence of a FIM/SCIM III crosswalk. Charlifue et al.[58] discuss the advantages of harmonizing databases in SCI to promote "new scientific discoveries and faster access to treatment interventions" and suggest that harmonization will enable meta-analysis, data pooling and advanced statistical approaches. However, prior to harmonizing databases, a mechanism is needed to compare data between them. Creating a FIM/SCIM III crosswalk or "link" will allow comparisons of functional

outcomes between databases and enable a broader understanding of published literature, in which FIM or SCIM III were used. A summary of these two major databases and how they have contributed to SCI research is provided below, to highlight the potential impact of a FIM/SCIM III crosswalk.

The US SCI Model Systems was established in 1970 and currently captures 6% of new SCI cases occurring every year in the US.[36] SCIMS is a network of rehabilitation centers that was established through an initiative of NIDILRR which maintains the largest longitudinal SCI database in the world. One of the original objectives of the SCIMS program was to develop a database to document the results of a system of care including rehabilitation outcomes and cost effectiveness.

This database, which has been maintained at the National SCI Statistical Center (NSCISC) since 1984, is a cohort database featuring standardized data, collected across the SCIMS centers, enabling comprehensive longitudinal studies of the natural history of people with SCI. Other objectives are to assess trends in service delivery, demographic characteristics, and treatment outcomes over time; to establish standards for rehabilitation outcomes based on results achieved by the Model Systems that can be used by other facilities for program evaluation; and to identify individuals who meet eligibility criteria for other studies who then might be invited to participate in those studies. As of September, 2019, the SCIMS National Database contains data on 34,130 Form I (initial acute and rehabilitation) participants, and 124,188 Form II (longitudinal follow-up) records among 28,463 participants. The longest follow-up is 45 years' post-injury. Extensive demographic and quality of life data are collected in addition to the International Standards for the Neurological Classification of SCI (ISNCSCI) and functional outcomes data, such as FIM (FIM data collection for long term acute care facilities ended in early 2019 due to the transition to CARE TOOL, but continues in rehabilitation facilities – discussed in Section 2.2.1).

The EMSCI network was founded in 2001 and currently consists of 16 active centers in 6 countries collecting data. The purpose was to establish a multicenter basis for future therapeutic

interventions in human SCI and to understand natural recovery and development of advanced clinical trial protocols. Individuals with traumatic SCI are assessed acutely (< 2 weeks after SCI) and at 4, 12, 24, and 48 weeks after SCI. The EMSCI assessments currently consist of the core data sets: neurological (ISNCSCI), function- specific COAs (locomotion and upper extremity function) and ADLs (SCIM III). Additional impairment assessments of neurophysiology and pain are collected. As of March 2019, over 4500 patients with spinal cord injury were included in the EMSCI database.

Data from both databases have been used extensively in research studies. Select examples from SCIMS which incorporated FIM, include studies on the relationship between acquired infection and long term recovery[59], assessing the value of FIM in predicting in economic burden[60], and the impact of age on SCI recovery.[61] Other SCIMS publications have focused on neurological recovery following SCI[62,63] with a fairly recent (2016) study examining neurological and functional recovery (FIM) in thoracic spinal cord injury.[64] Recent studies from the EMSCI data base in which SCIM III was used include research on predictors of functional outcomes[65], development of a linearized outcome measure using data elements from SCIM III and ISNCSCI[66], and the relationship between walking speed and community ambulation.[67] A number of studies using EMSCI data examined neurological recovery in thoracic [68] and cervical SCI.[69] Kramer et al.[70] and Steeves et al. [71]specifically looked at the relationship between motor recovery and function, using SCIM III, and from these data suggested that a two motor level improvement in neurological level of injury for individuals with complete, cervical injuries, confers a clinically meaningful change and further suggested a two level change as a clinical trial endpoint. This endpoint was derived from EMSCI data, but has been applied in North American based studies. Although this endpoint may be applicable for complete cervical SCI in differing systems of care, a better understanding of the relationship between FIM (primarily used in the US) and SCIM III (primarily used in Europe), would strengthen this recommendation. As noted in the introduction, the inpatient lengths of stay and thus

rehabilitation focus differs between these two geographic regions, which may impact functional recovery.

Both EMSCI, SCIMS and other databases have significantly contributed to our understanding of natural recovery, comorbidities associated with SCI and the long term impact of SCI. Creating a FIM/SCIM III crosswalk is a key step to leveraging data from different datasets to allow comparisons between datasets and published research findings.

2.4 A Brief Comparison of Classical Test Theory and Item Response Theory

Prior to describing linking methods, assessment of crosswalk strength and applications in health care, a brief comparison of classical test theory (CTT) and item response theory (IRT) is provided, as methods are described below using both approaches. The theoretical construct between CTT and IRT methods is quite different. Simplistically, CTT focuses on the test, where IRT focuses on test items.

CTT is based on the concept that an observed score is composed of the true score and error, and assumes that each item is equally difficult and that error is the same for all examinees. Thus, examinee and test characteristics cannot be separated. Test assessment in CTT focuses on classical psychometrics such as inter- and intra-rater reliability, predictive value, sensitivity and specificity. Typically, a longer test increases reliability, but one is unable to readily determine if individual test items add new information. The test properties are sample dependent, meaning that that error will vary in different samples. When comparing examinee ability, the same test (or a parallel test) must be used. In CTT, the test is therefore sample- and test-dependent. Test scores are often ordinal, meaning that the difference in difficulty between scores is not necessarily the same. An example is the assessment of muscle contraction strength (i.e. segmental motor score), where a difference between a score of “0” (total paralysis), and “1” (palpable or visible contraction), may not be the same as a change in score from “2” (active movement, full range of motion ROM) with gravity eliminated)) to “3” (active movement, full

ROM against gravity) (Table 4). CTT- based statistics often assume data are continuous, which can be an over-simplification, making their application to ordinal data problematic. The advantages of CTT approaches are that they are the most common approach to test development and assessment, as they are generally simpler and easy to understand. Disadvantages are the simplistic and often erroneous (when the incorrect statistical approach is used) statistical consideration of test scores.

Table 4: Muscle function grading

0 = Total paralysis 1 = Palpable or visible contraction 2 = Active movement, full range of motion (ROM) with gravity eliminated 3 = Active movement, full ROM against gravity 4 = Active movement, full ROM against gravity and moderate resistance in a muscle specific position 5 = (Normal) active movement, full ROM against gravity and full resistance in a functional muscle position expected from an otherwise unimpaired person

IRT is a more sophisticated, granular and multi-faceted approach, and focuses on test items and person ability. It is said to be test- and sample-invariant, such that the choice of test instrument is irrelevant. For example, if a person's height is measured with a yard stick in inches vs. a plumb line in centimeters, the person's height is the same, regardless of the instrument used to measure it. Also, it should not matter from which sample you draw data. IRT is based on the latent trait model, in that individual test items represent a single underlying trait or construct, which may not be observed directly, but anchors all the test items. This approach considers individual test item difficulty and person ability and determines the probability of a correct response and does not assume that all items are equally difficult. Redundant test items can be identified and removed from a test, as can items representing a different underlying construct. IRT uses a log transformation, which places all items on a common, linear scale eliminating the challenges with ordinal data. Thus the distance between test item #4 and #5 may be different than item #6 and #7, but the difference is quantifiable. Advantages of IRT are

that it considers more facets of a test or measure than CTT, and is thus a useful tool in the development and assessment of COAs. Disadvantages are that it is complex and harder to understand and the model itself can introduce error. A more detailed discussion of IRT and Rasch (the type of IRT used in this analysis) is provided in Section 2.9.4.

2.5 Overview of Linking Methods

An overview of linking also has been provided in Section 1.3. In this section, we will briefly describe linking methods, linking designs and how a crosswalk is assessed and validated. In the following sections, a more comprehensive description of each method is provided, followed by applications of that method in health care research.

Three conceptually different methods are used for linking, with different strengths and assumptions. Expert panel linking is a method whereby experts in the field establish equivalency for similar items and scores across instruments based on their expertise and experience. This is the "simplest" method, but is rarely used, presumably due to the inherent bias in this process and the development of more advanced methods. Equipercentile linking is a commonly used alignment method, rooted in classical test theory (CTT), in which a crosswalk is developed based on aligning total score distributions and rank ordering both total scores. Linking methods using Rash analysis co-calibrates items on a common linear scale and creates a crosswalk based on item difficulty, related to the underlying construct measured by the COAs in a more sophisticated and complex approach than CTT.

2.6 Data Collection Designs for Linking Studies

The most common data collection design for linking of COAs is the single group design or common person linking. This is the strongest design, in which both instruments are given to the same individuals (assessors may differ). This design allows for easier validation compared to other methods, as actual scores and linked scores are obtained on the same participants. Although this is the most

straightforward linking design, one disadvantage is the possibility of order effects, where the relationship between the COAs may be impacted by the order in which they are given. Commonly, retrospective data are used for crosslinking in which the order may not have been randomly assigned.

There are several alternative designs; the most commonly used are the equivalent and non-equivalent group design. In the equivalent group design, two sample groups from a common population are given both tests. This design eliminates order effect and requires less time for individual participants (as they only take one exam) but requires a larger sample size.

The common item non-equivalent group design is when different tests, with a sub-set of common items, are given to samples from two different groups of participants. The groups are not equivalent and differences in distribution of scores may be related to examinee groups or test differences.

2.7 Crosswalk Assessment

After a crosswalk is created, the strength of the crosswalk should be assessed and ideally validated in a separate sample population. The criteria for assessment are dependent on the intended purpose of the crosswalk. In the SAT/ACT linking examples discussed above, the crosswalk is used at the level of the individual to determine if scores on one exam are equivalent to scores on another exam. This information is used, in part, to determine college admission, thus it is important the crosswalk is assessed fully and validated at the level of individual scores. At the group level, a crosswalk may be used to compare a previous research finding using one test to another research finding using a second test, or for meta- analyses. Dorans[72] suggests evaluating the similarity of the constructs between measures, assessing the strength of the empirical relationship between measures and considering population invariance.[8] A number of methods have been used to assess crosswalk strength:

- **A correlation of ≥ 0.866**

The correlation coefficient assesses the strength of the relationship between measures. Although many different subjective cut-off values for correlation coefficients exist, Dorans[8] has suggested a minimally acceptable correlation coefficient specifically for linking of 0.866. This cut-off is in the strong to very strong range using the criteria in Section 2.2.1 of Schobert et al.[29] Various degrees of uncertainty can be calculated for a given a correlation (see formula in Appendix B) and Dorans[8] has justified this value based on the amount of uncertainty for a crosswalk to predict a person's score. With a correlation of 0.866, the uncertainty predicting a person's score based on the crosswalk is reduced by 50%. If uncertainly cannot be reduced by at least 50%, Dorans[8] suggests the prediction value of the crosswalk would be below an acceptable range. This recommendation is based on the importance of the crosswalk and whether individual scores are linked. In the SAT/ACT example, individual scores are linked and impact a candidate's admission to a given university. As noted by both Dorans[8] and Choi et al.[73], more relaxed criteria (0.75-0.80) may be used in situations where the outcome is less impactful and/or for group comparisons.

- **Population invariance between actual and cross-walked scores**

The difference in actual and converted scores should be similar between population sub-groups, which implies similar constructs between COAs. Common sub-group comparisons include males and females and age groupings. For example, the difference in actual and cross-walked scores between males and females should be similar. If they are not, one can consider different crosswalks for different sub-groups. A common method for assessing sub-group invariance is the root mean square difference. Two different "cut-offs" for acceptable sub-group invariance have been derived from a publication by Dorans and Holland[9], 0.8[74] and 0.11.[75] These numbers are similar to effect size and are interpreted as

such, wherein a lower value indicates a smaller difference between groups, and values lower than 0.10 are considered small.[76]

- **A low percentage of point differences between actual and cross-walked scores.**

Using this criterion, clinically relevant or statistically derived (from standard deviation) point differences can be identified and compared. A variety of methods have been used to examine point differences between actual and cross-walked scores. In some cases, reporting the percent of score differences (e.g. 1,2 or 5 points)[77-79], mean and median score differences[80], or score differences based on standard deviations were presented.[81]Others reported score differences related to changes in the amount of assistance required at a given score (e.g. 5 and 10 points for FIM) and used a 75% threshold as the cut-off for acceptable score differences.[82] In many cases, no justification for score differences or cut-offs was provided. Using MCID as a threshold is an appealing option as intuitively, a crosswalk with error that exceeds MCID would be less useful.

- **Similar score distributions between actual and cross-walked scores**

The first four moments (mean, standard deviation, skewness and kurtosis) should be similar between actual and converted scores.

- **Small effect sizes between actual and converted data**

Effect size will provide an assessment of the magnitude of the difference between the actual and converted scores. Ideally the difference is small. Cohen's criteria for effect size are 0.8 = large, 0.5 = medium, and 0.2 =small.[76]

- **Distribution of differences between actual and cross-walked scores**

Correlation coefficient reflect the strength of the relationship between two measures (do values on one measure increase when values on the second measure increase?) but correlation does not

necessarily indicate agreement between measures. Bland-Altman plots can be used to evaluate agreement between measures and where in the score distribution differences may exist.[83] A scatter plot or Bland-Altman plot can look for systematic differences between raw and crosswalked scores.[84-86]

Additional work by Byers[87], Tulsky et al.[81], Noonan et al.[75], Wang [82]and as noted by Ketchum et al.[88] considered criteria for linking individual vs. group level data. Criteria between studies differed in what was considered essential for a crosswalk at the individual level. Criteria considered for individual cross-walking in at least one of these studies was a correlation ≥ 0.866, point differences between actual and cross-walked scores, and score distributions.

2.8 Crosswalk Validation

Ideally, after initial crosswalk development, the crosswalk should be validated in a second independent sample. Where a separate, independent sample is not available and there is a sufficiently large subject population, the original sample can be split into a development and validation sample. The crosswalk is created and assessed in the development sample and is then applied and assessed in a second sample for validation.

2.9 Crosswalk Development Methods

2.9.1 Expert Panel Linking

The expert panel linking method is when experts in the field of interest examine both measures and determine which items and scores can be linked. This method can be challenging when items and scoring differ between instruments. Eliminating items that do not match may result in loss of valuable data that contributes to the instrument's construct and may impact instrument psychometrics.[89] As noted previously, both this method and equipercentile linking are sample dependent, meaning the outcome is dependent on the sample from which the data were drawn.

2.9.1.1 Examples of Expert Panel Linking in Health Care Applications

Very few examples of expert panel equilibration exist in the literature, presumably because this method is subjective and newer, statistically based methods such as equipercentile and IRT methods have now been developed. One example is that of linking the FIM and the Minimum Data Set (MDS).[90] FIM is commonly used in the rehabilitation setting, while MDS is used in long term care facilities. Prior to 1997, no link existed between these instruments, thus it was difficult to track progress over time or to compare individual characteristics and outcomes. In this first FIM/MDS linking study, Williams et al.[90] asked seven experts to identify common FIM/MDS items and as scoring differs between the two instruments, to link scores as well. Scoring links were intended to include "the most and least limited individuals" within a particular score range, which was important as scoring differed between instruments. For scores in which multiple scores on one instrument corresponded to a single score on another, the average was used. For example, if a FIM of 3 and 4 corresponded to an MDS of 3, the MDS level was defined as 3.5. Agreement was obtained when 2/7 panel members agreed. For most items, the majority of panel members agreed. In a second method of rescaling, prospectively collected FIM/MDS data (173 paired data sets) was split into a development and validation data set. In the development dataset, the mean of the observed FIM levels for each for each MDS level were calculated. The re-scaled MDS items for expert derived FIM scores, are referred to as PseudoFIM(E), and the MDS values based on observed means, PseudoFIM(O). PseudoFIM(E) and PseudoFIM(O) values were then applied in the validation dataset and compared to actual FIM values. Absolute difference in mean item scores between actual FIM and PseudoFIM(E) were calculated, with 7/12 items demonstrating no significant difference. Findings for the FIM to PseudoFIM(O) were similar, although PseudoFIM(O) was more closely aligned with actual scores (8/12 items). T tests revealed no differences in subcales between FIM and PseudoFIM(E) scores.

2.9.2 Overview of Observed Score Linking Methods

Several types of linking can be conducted based on classical test theory. Mean linking can be used when the difference in difficulty between scales is the same throughout the scale (Figure 1). In this case, the differences in scores due to item difficulty are equivalent at the upper, middle and lower ends of the scale. While the means between the scales may differ, the scale units (or difference between scores) do not, such as when the difference in the means is two points throughout the scale. When graphing both instruments, the slope of the line will be the same for both measures, and the distance between both lines will be the same at all points along the line, although the lines may lie at different points on the X axis. In this case, the same constant can be added to all scores on Test A to achieve equivalence on Test B.

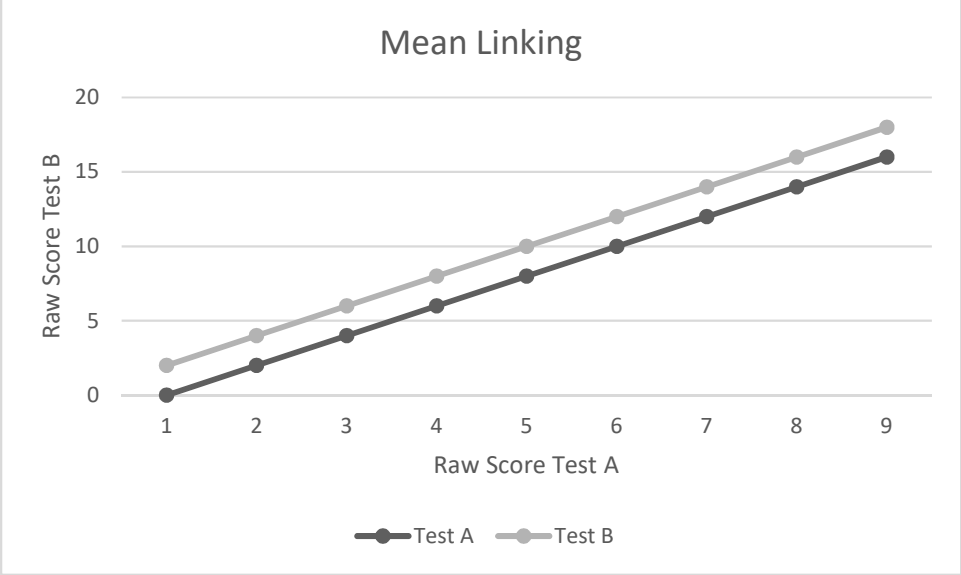


Figure 1: Graphical example of mean linking

Hypothetical scores for Test A and Test B are graphed, where the difference in means is the same throughout the scale.

Linear linking accounts for differences in difficulty along the test scores (Figure 2). For example, Test A may vary in difficulty relative to Test B at the higher versus lower ends of the scale. A score of "0" on both tests may be equivalent in difficulty, such as the case where a score of "0" means an individual cannot complete the task (item difficulty is similar), and a score of "7" on Test A may indicate an individual can climb a flights of stairs with no handrail, whereas the same score on Test B refers to climbing three stairs while using a handrail. Thus, both means and standard deviations may differ. This is illustrated when graphing scores on both instruments, where the slope of the line differs and the distance between the lines varies at the upper or lower ends of the line. Linear linking accounts for this by matching score distribution and standard deviation. Use of this method does not allow for a non-linear relationship between measures (i.e. most ordinal data).

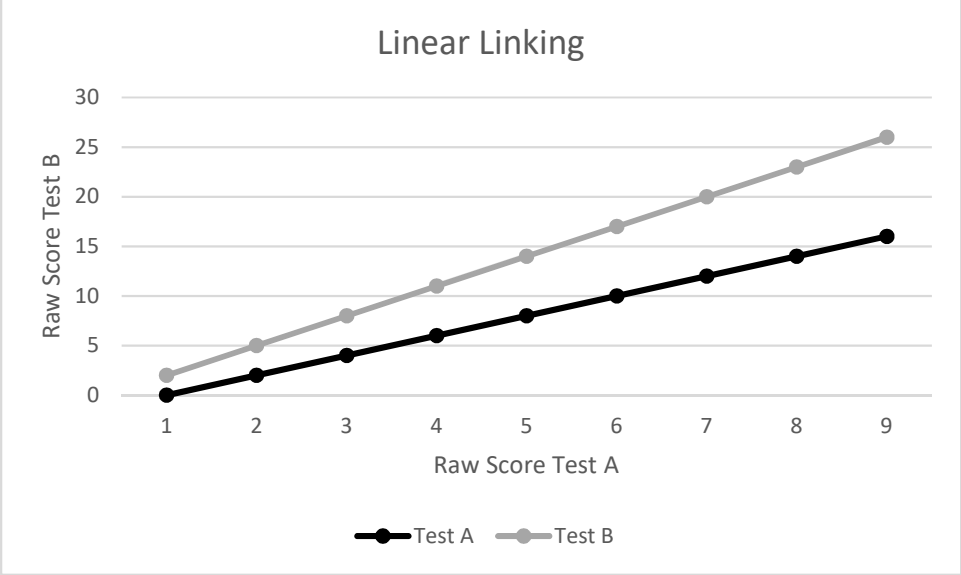


Figure 2: Graphical example of linear linking

Hypothetical scores for Test A and Test B are graphed, where the means differ linearly throughout the scale.

Equipercentile linking allows for a curved (non-linear) relationship between measures such that Test X may be more difficult at high and low scores, but less difficult in mid-range scores. In health sciences, this is the most common type of linking based on classical test theory, as it is the most flexible and is the most applicable in terms of linking instruments such as FIM and SCIM III, where difficulty between tests may vary throughout the score range. Also, with mean and linear linking, due to addition of a constant to the score range, equated scores can occur which are outside the possible range of scores, which is not the case in equipercentile linking.

2.9.3 Equipercentile Linking

Equipercentile linking matches the percentile rank of examinees achieving a given score on Test A, with the same percentile rank of a given score on Test B. The crosswalk is developed by identifying scores on Test A with the same percentile rank as scores on Test B. For example, if the 50th percentile of a score on Test A is 35, and on Test B is 42, these two scores are lined up and the link for a score of 35 on Test A is a score of 42 on Test B. Matching test scores to percentile rank is conducted for all test scores. Tests scores for Individuals who fall below the 50th percentile on Test A are not necessarily the same individuals who fall below the 50th percentile on Test B. A crosswalk table may then be created linking scores on Test A with scores on Test B.

A simplistic example of a four-point scale, similar to that provided by Kolen and Brennan[91] is provided below. In Table 5, for Test A, A = test score, proportion is the proportion of examinees with a given score, cumulative proportion is the cumulative proportion at or below a given score of A, with percentile rank in the final column. Figure 3 graphs raw score and percentile rank for Test A and Test B. To find a linked score between tests, a vertical line is drawn and the equivalent score at a given percentile is identified. For example, to determine the Test A score for a score of three on Test B, a vertical line from a raw Test A score of three which falls at the 60th percentile rank. A horizontal line is

then drawn to Test A at the same percentile rank, and a vertical line is drawn to the equivalent score, which in this case is 3.75. Linking tables can then be created matching equivalent scores for a given percentile rank.

Table 5: Hypothetical table for equipercentile linking

Test A Score	Proportion	Cumulative proportion	Percentile rank		Test B Score	Proportion	Cumulative proportion	Percentile rank
0	0.1	0.1	5		0	0.2	0.2	15
1	0.2	0.3	15		1	0.3	0.5	35
2	0.2	0.5	35		2	0.2	0.7	55
3	0.3	0.8	60		3	0.2	0.9	75
4	0.2	1	90		4	0.1	1	90

proportion = the proportion of examinees at each score, cumulative proportion = the proportion of examinees at or below the score, percentile rank= the percentage of examinees at or below the score

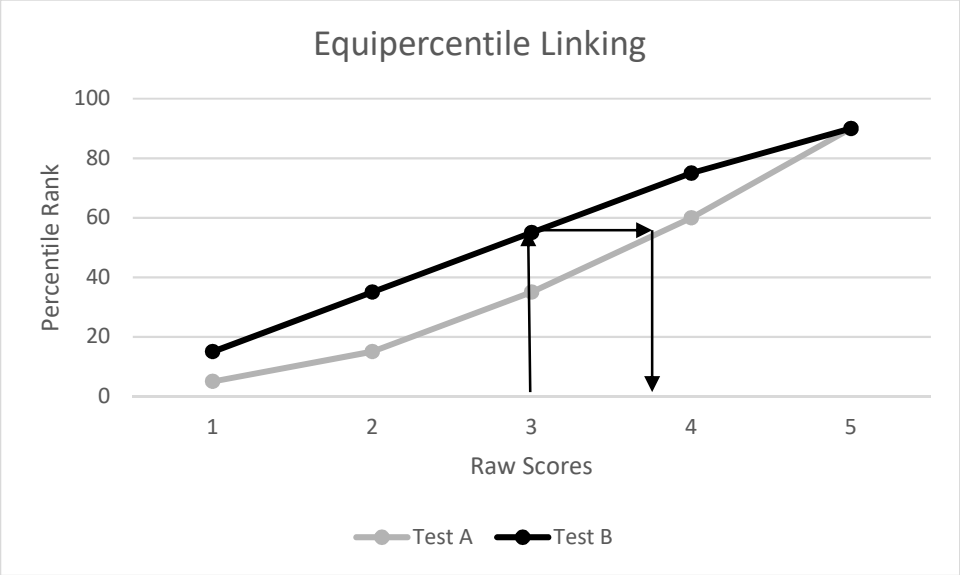


Figure 3: Graphical example of equipercentile linking

Hypothetical scores for Test A and B are graphed with percentile rank, where the means between tests may differ non-linearly throughout the scale. In this example, a score of 3 on Test B is equivalent to a score of 3.75 on Test A.

As sample percentiles and percentile ranks are used to create an equipercentile linking function, sampling error in which some scores are overrepresented while others are underrepresented or perhaps not represented at all, may be present. Sampling error will be greater in smaller samples. The need for smoothing to decrease sampling error can be assessed by graphing the equipercentile relationship between the two tests. If the relationship appears even, smoothing is not required. If the relationship is jagged, smoothing may improve the linking relationship. The two approaches used for smoothing are pre-smoothing in which the score distributions for each form are smoothed, or post-smoothing where equipercentile equivalents are smoothed.

2.9.3.1 Examples of Equipercentile Linking in Health Care Applications

Equipercentile linking is frequently used in health sciences, particularly in psychology. A number of studies used common person, equipercentile linking with log linear smoothing to create crosswalk tables for measures of cognitive function. The focus in many of these studies was on validation of new measures. Crosswalk tables were created to enable comparisons in longitudinal studies. As the focus of this dissertation is on cross-walking, we have not focused on the strengths and challenges regarding traditional COA validation in these publications, but on equipercentile linking.

In a series of publications linking schizophrenia measures, authors (Leucht et al.[92-95] Levine et al.[96,97], Schennach-Wolff et al.[98]) linked a clinical global impression (CGI) scale of severity and improvement, with outcome measures assessing severity of illness based on specific psychological symptoms, assessed by interview. As the global impression scale was more understandable to clinicians than the specific symptom scales, the intent was to provide context for clinicians to better understand the symptom specific scales and to identify cut-points that could be used in clinical trials. The majority of these studies re-analyzed clinical trial data from drug studies, while two trials used hospital admission data, one prospective and one retrospective, in order to sample a broader population. Total scores of

the clinical global impression scale for severity and improvement were linked with symptom specific scales at a variety of time points. In some studies, percent and absolute change on both instruments were also linked. As crosswalk development is simply about the relationship between instruments at a given time point, when clinical data are used, data from the active and placebo/comparator arms can be used.

In terms of assessing key crosswalk criteria such as similarity of content, adequate correlations between the linked measure and population invariance, the only assessment was correlation coefficients. Often, it was unclear if the correlations presented were between raw scores on the instruments (assessing the relationship between instrument) or raw and crosswalked scores, (assessing crosswalk strength). Graphs illustrating linking functions were presented, but no crosswalk tables were provided. In terms of content, no formal analysis of content was noted, but content and intent of the scales are quite different. For example, The CGI rates a clinician's impression of severity of illness and severity based on clinician experience with this population. It uses a seven-point scale with subjective severity items ranging from "normal – not at all ill, symptoms of the disorder not present past seven days" to among the most extremely ill patients – "pathology drastically interferes in many life functions; may be hospitalized". The CGI improvement scale is similar (very much improved to very much worse). The symptom severity scales are rated on specific symptoms such as delusions, grandiosity, depression, etc. based on a 30-45-minute interview. Thus, key criteria for content similarity was not met.

Spearman correlation coefficients between the linked measures ranged from 0.41 – 0.75, which do not meet the criteria for individual comparisons suggested by Dorans (0.86)[8] and only met the lower end of the relaxed criteria for group comparisons suggested by Choi (0.75-0.80).[73] No sub-group invariance analyses were conducted and although these studies were replicated in numerous data sets with similar findings, no validation studies occurred. The reason for the below threshold linking

correlations may be related to the different intent and content of the measures. Additionally, Leucht et al.[93] noted that "psychometric characteristics (of the CGI) are not well defined". Low reliability can impact the correlation between instruments. One intended purpose was to make the symptom scales more relatable to clinicians based on verbiage from the CGI. Although this purpose could be considered a "low bar" in terms of linking, the consistently low correlations may still make this connection unclear or erroneous.

Researchers also linked two measures of social and occupational function and linked these measures with a psychological symptom-based measure and a CGI severity scale.[99] Correlations between the occupational and social measures were stronger (0.86-0.93) than those between occupation and social measures and symptom based measures or CGI measures (0.57-0.71). Stronger correlations were likely due to the similarity of the constructs assessed by the social and occupational measures, which meet the minimum criteria suggested by Dorans.[8]

In another series of nine studies equipercentile linking was used for cognitive assessment in dementia, Alzheimer's disease, and Parkinson's disease.[77-79,100-105] In all but one of these studies, log-linear pre-smoothing was applied. Correlation coefficients (reported in 7/9 studies) ranged from 0.66 – 0.94, with one study[101] exceeding the recommend 0.866 threshold for individual score comparisons and 6/7 above the more relaxed threshold of 0.75-0.80 for group comparisons. Similar to linking of schizophrenia measures, in most cases construct similarity was not formally discussed. Contrary to the schizophrenia studies, two of these studies validated prior crosswalks in separate samples[78,79], and in one study validation occurred within the same study by splitting the sample into a development (70%) and validation (30%) sample.[77] Validation criteria used in these studies entailed a comparison on point differences (none, one, two or more) between actual and linked scores in the validation dataset. Lawton et al.[79] also calculated the difference between actual and linked scores of the mean, standard deviation,

median and interquartile range as well as the root mean squared error, with smaller values indicating smaller differences.

In a recent study assessing two general anxiety measures for individuals with an implantable cardioverter, equipercentile methods were employed with log linear smoothing.[84] A crosswalk table was provided, as well as correlations between raw and crosswalked scores (0.75). Investigators also examined the way in which crosswalked scores performed in terms of clinical cut-off scores. Confirmatory factor analysis was used to better understand the two measures but authors did not link those findings to the decision to create a single crosswalk. Other than cut-off scores and a scatter plot comparing raw to crosswalked scores, no secondary criteria were examined and validation of the crosswalk did not occur. Authors concluded that cutoff scores between the measures were not comparable.

Two studies are presented below that assess various aspects of physical function, collected by questionnaire. In 2017 Ghomrawi et al.[106] created a crosswalk between two self-administered lower extremity activity scales for individuals with total hip or total knee replacement (n=767). Two-way crosswalk tables were created. Crosswalk validity was assessed by comparing means for actual and converted scores. Standard response mean (SRM), to assess responsiveness to change, was compared between actual and converted scores. Differences in receiver operating characteristics (ROC), to discriminate functional thresholds between actual and converted scores were also compared. Means, SRM and ROC did not differ between actual and converted scores. No correlations were provided.

In a multiple sclerosis (MS) study, Noonan et al.[75] took a slightly different approach. The objective of this study was to identify the appropriate crosswalking method for measures of fatigue (PROMIS Fatigue Short Form and Modified Fatigue Impact Scale- MFIS), based on criteria established by Dorans[72], and then create and validate the crosswalk. Noonan et al.[75] evaluated the similarity of

constructs by assessing content of each instrument as well as conducting factor analysis. As some content differences were noted and confirmatory factor analysis suggested the MFIS may not be unidimensional, authors then ran exploratory factor analysis (EFA). EFA provided support for a single dimension in both scales, based on established criteria. In addition to comparing score distributions for both measures and calculating correlation coefficients for linked scores, this was the first study to assess an additional fundamental criterion for linked scores, population invariance. Population invariance was evaluated by calculating the standard mean differences (SMD) for age categories, type of MS, and duration of MS. SMD was calculated by subtracting the mean score of group 1 (in this case relapsing and remitting MS) from the mean score of group 2 (other types of MS) and dividing by the total group standard deviation (SD). Criteria of a difference less than 0.11 was used, with reference to Doran's work.[8]

In this study, the development data set was obtained at a single time point in data collection, while a later time point was used for validation. Crosswalk validation occurred by projecting total scores of one measure based on actual scores on the other measure and vice versa. Deviations between actual and converted scores were then calculated. The values of these deviations were assessed over the score range to determine differences in the high vs. the low range of the scales.

As sample size can significantly impact deviations, Noonan et al.[75] also conducted bootstrapping with samples of various sizes to determine sample sizes with acceptable variability. Estimated crosswalk-based means for the random samples were compared with observed sample means. They estimated a sample size of ≥ 150 was required for group comparisons, due to deviations well above two points in smaller samples. The authors noted that although the correlation approached 0.866, results did not support using the crosswalk for individual level data due to the sample sizes required for comparison. They also noted the differences in how the underlying construct was assessed in both measures and

stated that individual cross-walked scores were not interchangeable. The study was the most rigorous study reviewed, whereby crosswalk strength was assessed based on recommendations by Dorans.[72]

These initial studies significantly contributed to the literature and maturation of equipercentile linking for crosswalk development. Many of the early studies included limited assessment and validation of the crosswalk, thus their utility is questionable. Discussions of content and construct within and between instruments were rare and of those studies reporting correlation coefficients, only one met the threshold suggested by Dorans.[72] Additional criteria of sub-group invariance and comparisons of score distributions were rare. Where point differences between raw and crosswalk scores were compared, the rationale for the chosen point differences was not provided. Methods similar to those used by Noonan[75] should be considered in future studies and were used in this dissertation.

Only two of these studies evaluate physical function, and none evaluated physical performance. No equipercentile linking studies in SCI were identified in the literature.

2.9.4 Background on Item Response Theory and Rasch Analysis

As described above, Item response theory (IRT) is a family of modern statistical approaches that often are used in test development, item banking and test linking. IRT is a more sophisticated approach than traditional classical test theory in that classical test theory focuses on test scores, whereas IRT considers individual test items and person ability. IRT is a probabilistic model (the probability of a correct response), based on a person's ability and item difficulty.

Within IRT, there are different models, reflecting the number of parameters within the model. In three parameter IRT the following items are considered: 1) item difficulty; 2) item discrimination, which reflects the degree to which an item discriminates between persons of different abilities (e.g. a person with lower ability has a smaller chance of responding correctly) related to the underlying dimension; and

3) guessing, as with a multiple choice exam. The two parameter IRT model does not consider the effect of guessing, and the one parameter model does not consider item discrimination. Rasch analysis is considered a one parameter IRT model, although some of the fundamental concepts differ from IRT. In the Rasch model, the data are fit to the model whereas with IRT, the model is fit to the data. With Rasch, the degree to which the data do not fit the model is then analyzed. For items that do not fit the model (e.g., assess a different dimension), items can be removed until better model fit is achieved in an iterative process. IRT and Rasch are both used in test linking. In a study by ten Klooster et al.[107], one, two and multi- parameter linking methods were used and the resulting crosswalks compared. In this case, the one parameter (Rasch) cross-walk, performed similarly to the more complex models.

Core tenets of IRT include unidimensionality and local independence. Unidimensionality means that a test measures a single underlying trait such that differences in responses reflect item difficulty and person ability, vs. a different skill or trait. Local independence means test items are independent, thus responses are related to person ability. The degree to which these traits hold is assessed with Rasch fit statistics, which will be expanded on below.

The Rasch equation provides an understanding of the factors considered in Rasch analysis: Log [Pnijk/Pnijk-1] = Bn –Di – Cj – Fk (also presented in Appendix B). As summarized in Velozo[108], simplifying the equation, the left side is the probability of a person passing a test item, divided by the probability of not passing a test item. It also represents the probability of receiving a particular rating, versus the rating above or below that rating. This is presented as the log value which converts ordinal scores to more linear interval scores.

The right side of the equation includes the interaction of person ability (Bn), item difficulty (Di), rater severity (Cj) and the difficulty of rating step k, relative to step 1-k (Fk). Bn-Di is a core component of Rasch measurement and relates to the concept of scale-free measurement. This means that person

ability is invariant and that the choice of assessment scale is theoretically irrelevant, so that person ability can be compared, regardless of the test. Rasch also accounts for item difficulty and rater severity, which are not considered in CTT. Fk relates to the structure of the rating scale, such as the "ordinality" of the scale and accounts for differences in difficulty between test items.

Operationally, scale-free measurement occurs by co-calibrating items (in this case items from two COAs) and persons on a common, linear scale as illustrated in Figure 4. Co-calibration allows for assessment of measurement properties such as dimensionality, item and person fit, item and person separation, and item hierarchy.

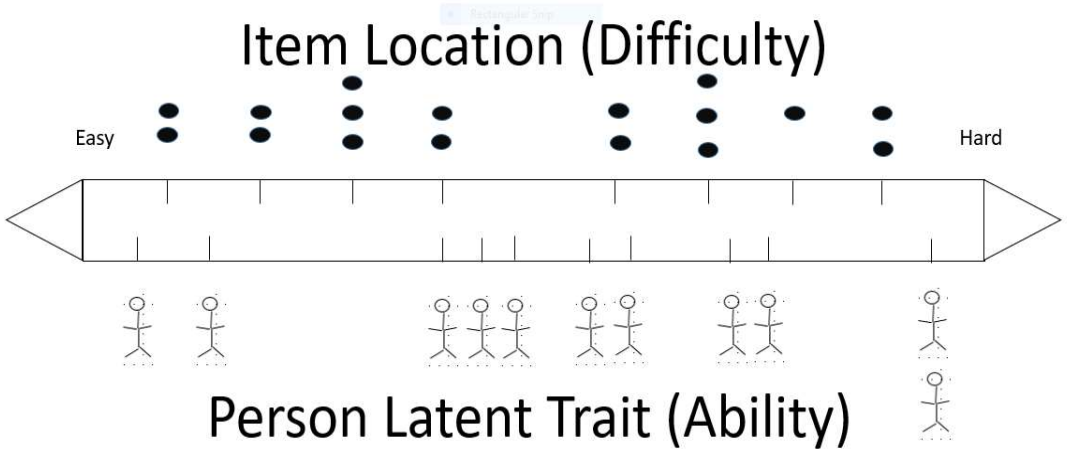


Figure 4: Illustration of co-locating item difficulty and person ability on a linear scale

Easier items and low person ability are on the left side of the ruler, with high item difficulty and person ability on the right side of the ruler.

Unidimensionality, or measurement of a single underlying construct or latent trait is a requirement and assumption in Rasch analysis (and linking), which can be assessed in several ways.

To understand the terminology, a bit of background is required. Eigenvalue is a term used in relation to principal components analysis (PCA) in Rasch, and other mathematical applications. PCA converts a set of observations of possibly correlated values into a smaller number of uncorrelated

variables called principal components. The eigenvalue is a numerical representation of the variance explained by the principal components analysis.

In Winsteps (a Rasch software program), an eigenvalue of the first contrast <2.0 indicates random noise, whereas an eigenvalue > 2.0 is indicative of the number of values that may represent a separate construct.[109] Dimensionality can also be assessed in factor analysis (described below) which is often completed prior to a Rasch analysis of dimensionality.

Item infit and outfit assess the "fit" of particular test items to the Rasch model, and quantifies discrepancies. This assessment is primarily used to identify items that do not fit the unidimensional model, and can also be used to determine if items are redundant and should be discarded. Infit is more sensitive to inlier-sensitive fit (items with difficulty close to the person ability, i.e. typical values), outfit is more sensitive to items with difficulty far from person ability (i.e. outliers). Outfit and infit are calculated from mean square residual summary statistics, where a 1.0 indicates perfect agreement. Means less than 1.0 indicate "overfit" (i.e. redundancy), where means greater than 1.0 indicate "underfit". Mean square fit shows the size of the randomness in the data. For example, a mean square of 1.4 indicates that there is 40% more randomness in the data than the model. Different criteria have been used; Wright and Linacre suggest a reasonable range for infit and outfit is 0.5-1.7 for assessments obtained via clinical observation.[110]

Person fit assesses abnormal or improbable response patterns. In the case of a common person linking, person fit may indicate person data points with unlikely values that may be unexpected. Wright and Linacre[110] note that as items are encountered by many persons (and are stable), but persons encounter few items (and are less stable) , therefore, a degree of person misfit is expected. Thus approaches to person misfit are more relaxed and a few rogue values will not have a significant impact.

In one study, Malec et al.[111] used a cut-off of > 3.0 to exclude individuals with abnormal response patterns.

Person and item separation can be calculated in Rasch. Person separation identifies categories or strata, such as low or high functioning individuals. Low person separation (< 2, person reliability < 0.8) indicates the instrument may not be not sensitive enough to distinguish between different strata and additional test items may be needed. Item separation verifies the item difficulty. Low item separation (< 3 = high, medium, low item difficulties, item reliability < 0.9) suggests the sample is not large enough to confirm the item difficulty hierarchy of the instrument.[109]

Item hierarchy examines whether hierarchical items or scores are disordered. For example, in an item with a 1-5 scoring metric, where a score of one represents the easiest task and five represents the most difficult task, is three more difficult than two? If not, the test scores are considered to be disordered which may contribute to item misfit.

2.9.4.1 Factor Analysis

As noted above, dimensionality can be assessed in Rasch, and/or with FA. Unidimensionality is a requirement/assumption of Rasch and is also required for instrument linking. Factor analysis (FA) approaches dimensionality from a CTT perspective whereas Rasch assumes the latent trait model. As dimensionality is a continuum (not black or white), assessing it in multiple ways is warranted. FA often precedes and accompanies Rasch analysis.

FA is used to identify the strength of the relationship between items in an assessment and explores the underlying reasons why those items are correlated. FA does this by identifying underlying factors (also referred to as latent traits or dimensions), that may not be measured by the assessment, but contribute to the relationship between items. With FA one can see the strength of the relationships

between these items and based on which items correlate with which traits, may be able to infer why they are related.

Exploratory factor analysis (EFA) is used when the number of underlying traits is not known, whereas one may use confirmatory factor analysis (CFA), when the number of factors can be roughly identified and which variables will "load" or are affiliated with that factor thus, testing a specific hypothesis. EFA and/or CFA may be used prior to conducting Rasch analysis. FA may also inform Rasch analysis, in that if multiple dimensions are identified, multiple Rasch models may need to be created.

Although EFA is used when the number of underlying traits is not known, one needs to determine how many factors to assess. This is based on knowledge of the measures and the scree plot. Scree plots plot the number of potential factors against the eigenvalue. The first factor will always have the highest variance. The location of the "elbow", where a significant reduction in the eigenvalue occurs, is an indication of the number of factors to extract (Figure 5). Typically, multiple numbers of factors around the "elbow" are extracted and compared. Rotation is used to find the factors with the strongest item-total correlations. The choice of rotation method depends on the relationship between factors. Oblique rotation assumes that factors are correlated, while orthogonal assumes they are not.

A number of factors can be considered when determining the number of underlying traits in a COA. Noonan et al.[75] Reeve et al.[112] and the UCLA consulting group considered a combination of the following:

- Scree plot: The location of the "elbow" is indicative of the number of factors (Figure 5).[75,112]
- Magnitude of the eigenvalues: A minimum of 20% of the variability should occur in the first eigenvalue.[75,112]
- Ratio of the first and second factor: A ratio > 4 is ideal.[75,112]

- Total variance explained: In social sciences, the chosen number of factors should explain 60-80% of the total variance.[113]
- Pattern of factor loadings in the pattern matrix. Are items clearly loading on one factor or another or are multiple items loading on multiple factors?[75,112]
- The correlation between factors. High correlations indicate overlap between factors, where low correlations indicate distinct factors with minimal overlap.[112]

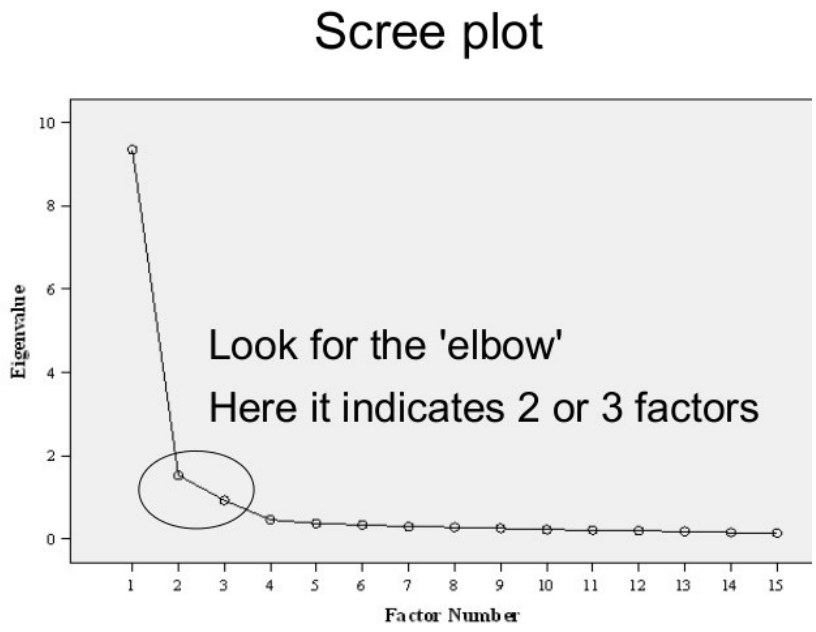


Figure 5: Sample scree plot

2.9.5 Rasch Analysis of FIM and SCIM

Now that the background on Rasch analysis has been provided, a summary of Rasch analyses of FIM and SCIM III is presented. While two Rasch analyses of SCIM III have been published[46,66], as of 2011 Nilsson and Tennant[114] provided a list of 50 publications where Rasch analyses of FIM have occurred in a variety of impairment groups, for different purposes. SCIM III and FIM Rasch analyses are briefly

summarized, predominantly in the SCI population, with some mixed impairment population publications where relevant.

Two early FIM Rasch analyses in mixed impairment publications by the same group[17,115]considered dimensionality across impairment groups using factor analyses and Rasch item misfit. Differences in admission vs. discharge FIM fit statistics and calibrations (in logits) were also examined as authors identified the need for FIM to function the same way at admission and discharge to ensure change was "true change" vs. instrument differences between time points. Authors concluded that FIM consists of two dimensions (or domains), cognitive and motor. After removing cognition, bladder management, bowel management and stairs and eating still demonstrated misfit at discharge, using criteria of 0.7-1.3 for item fit.[115] Authors noted that misfit is common at the most extreme test items (easiest and hardest) e.g. most difficult- stairs and easiest- eating, as these items are often less well-defined. Stair misfit may be due to a common score of "total assistance" when items were not observed, as a therapist may often choose not to test stairs due to safety concerns. Authors concluded that clinical importance of the misfit items took precedence over removing these items. Although disordered scoring was not examined in these publications, it is possible that if disorder was present, rescoring might solve item misfit. In a separate publication, PCA was used to examine dimensionality across impairment groups. SCI represented 6% of the sample (n=1,727) with a correlation of 0.68 with the first factor, -0.25 and -0.01 with the second and third factors respectively.[17] Although not reported in table format, authors noted misfit in bowel management, bladder management and stairs across all impairment groups.

Numerous SCI-specific FIM Rasch analyses have been conducted to develop a phone FIM[116] and computer adaptive test[117], to assess cross cultural validity[118], and to examine individual growth curves[119]. In the SCIRehab studies, Whiteneck et al.[120] used Rasch FIM linear transformations (on a 100-point scale)

for regression analyses. Score conversions for all motor items and subsets of motor items are presented in table format by Kozlowski and Heinemann.[119] Although Rasch analysis was used for different purposes, in those studies presenting item hierarchy, eating is consistently the easiest item, while stairs is the most difficult. Between studies, items with difficulty between eating and stairs may vary somewhat in terms of item difficulty and misfit.

SCIM III has also undergone Rasch analyses by the instrument developers[46] and as part of the development of a new linearized measure using data elements from SCIM III and motor ISNCSCI scores.[66] The Rasch analysis by Catz et al.[46] was conducted on sub-scales based on rehabilitation admission data. Item misfit was acceptable at the sub-scale level but eating, respiration, mobility outdoors and stairs demonstrated misfit using criteria of 0.8-1.4. Based on distribution maps organized by sub-groups, it can be inferred that feeding was the easiest item, while stairs was the most difficult. Disordered scoring thresholds were identified in respiration/sphincter management and mobility for all distances and stairs. A challenge of this Rasch analysis is that it was conducted at the sub-scale level (separate analysis for the three sub-scales), thus results may differ if approached at the individual item level.

As part of the development of the Spinal Cord Ability Ruler, Reed et al.[66] conducted Rasch analysis at the item level and to correct for item misfit collapsed highly variable scale scores in SCIM III to 4 levels. In their analysis grooming was slightly easier than feeding, while stairs were the most challenging item.

Common themes emerge from FIM and SCIM III Rasch analysis. Although item misfit and score disorder were noted in both measures, the COAs were not altered as a result of this information, in part due to the clinical utility of the items. Eating/feeding is generally the easiest item in both scales, while

stairs is the most difficult. These same items also demonstrated item misfit in both scales, perhaps due to less well described items at the extremes of both scales.

For both scales, although total score reliability from CTT was in a reasonable range, for individual items, reliability was low. Addressing disordered scoring for both instruments could improve item fit but if item fit remains unacceptable, removal of items could be warranted and might improve reliability for both instruments.

2.9.6 Rasch Co-calibration in Linking Functional Outcome Measures

In the first two studies linking observed measures of ADL function, Velozo et al.[89] and Wang et al.[82] created crosswalks between FIM and the MDS. As FIM typically is used in the inpatient rehabilitation setting and MDS in skilled nursing facilities, the goal was to enable assessment of function across the continuum of care. As noted above, a prior study created a crosswalk using expert panel linking.[90] This series of studies were the first to use Rasch analyses to create a crosswalk between these two measures. Rasch was chosen as the preferred linking method, as authors noted the challenges of expert panel equivalency include the lack of correspondence between all items and scores, necessitating dropping some items and scores, which may impact (improve or degrade) the instruments psychometric properties. In addition, a noted advantage of Rasch is that it is intended to be test and sample independent.

Common person linking using retrospective Veteran's Administration data was used in both Rasch studies. In the first study, FIM and MDS data were collected within 7 days, for any type of impairment, in a total sample size of 254 participants. Due to scoring differences, FIM and MDS data were re-scored. What was considered to be invalid data (data falling outside the 95% confidence interval on the identity line) were removed. Using the Rasch partial credit model, which allows for polytonomous responses with differing scaling between instruments, FIM and MDS items and rating scales were placed

on a common linear scale. FIM and MDS were then separately anchored to the common scale. A crosswalk table was then produced linking raw scores on FIM to raw scores on MDS. Crosswalk assessment included analysis of the psychometric properties of the co-calibration such as person and item level psychometrics, item fit and point measure correlations to assess unidimensionality and internal consistency. Item hierarchy was compared and a co-calibrated item map was presented, indicating item hierarchy for each scale relative to person measures. Correlations for raw scores (-0.81) and cross-walked measures (0.78) were calculated. Authors noted that that the crosswalk may be improved by decreasing the time between assessments (in the event differences in FIM and MDS score are related to actual change vs. crosswalk error) and /or the use of a more unidimensional instrument by removing ambulation and incontinence. Different item calibrations across groups using differential item functioning might also improve the crosswalk. Authors also noted that while item removal may improve the crosswalk, this may alter psychometrics of the original measures.

A follow-up study led by Wang et al.[82] was noted as a validation study, however, as Wang created and validated a second crosswalk from a different dataset, it is more appropriately characterized as a replication study.[82] In this study, a criterion of five days between assessments was used. The original data set was split into a development data set (n=654) and a validation dataset (n=1,476) and only three impairment groups were included (stroke, amputation, orthopedic impairment). The same Rasch procedure used by Velozo[89] was used to create the crosswalk, which was then tested in the development set at three different levels: 1) individual patient, 2) classification based on functional-related groups, and 3) by facility. Wang also removed seemingly “invalid” data. Individual level validation was assessed by comparing score distributions, the percent of 5- and 10-point differences and correlations between actual and cross-walked scores. At the classification level, the association (chi square) strength of the association (kappa) and percent of patients in the same

functional-related groups were assessed. Facility level validation consisted of comparison in score distributions between actual and cross-walked scores by facility, with paired t-tests to assess equivalence of mean scores between facilities.

Significant differences in score distributions (Wilcoxon signed rank test), and 5- and 10-point differences (33.7 and 56.9 % respectively), between actual and cross-walked FIM motor scores led authors to conclude the crosswalk could not be used for individual item level data. Correlations (Pearson) were 0.79, which the authors did not highlight as a concern.

These two studies are notable in that they were the first to use Rasch analysis to develop a crosswalk between two observed ADL assessments. Dimensionality, a key criterion for crosswalks, was assessed. Correlations also were assessed and, although the criteria suggested by Dorans[8] was achieved, authors did not use this "cut-off" as a key criterion to assess crosswalk strength, nor was sub-group invariance established. Differences in score distribution and point differences between actual and crosswalk scores of 5- and 10-point differences were assessed, although justification for these values was not provided. The replication study by Wang et al.[82] did not compare the derived crosswalk with that obtained by Velozo et al.[89], nor were crosswalk tables presented.

In comparing studies by Velozo et al.[89] and Wang et al.[82] to one another and the expert panel method of developing a FIM-MDS crosswalk used by Williams et al.[90], the actual vs. cross- walked score correlations for motor FIM were 0.81 (Spearman) and 0.72 (Pearson), a correlation for the whole scale was not provided. Velozo et al.[89] only reported the correlations between FIM and MDS raw scores and FIM and MDS crosswalk scores (-0.81 and 0.78 respectively – statistical test used was not described), but did not report the correlation between actual and cross-walked scores. Wang et al.[82] reported a Pearson correlation between actual and cross- walked motor scores of 0.79. Based on these data the Rasch method employed by Wang et al.[82] resulted in a slightly higher correlation coefficient (0.79) than the

expert panel method employed by Williams et al.[90] (0.72). However, as the datasets differed, one cannot conclude superiority of one method over another.

A later study in 2018 used Rasch analysis to link the FIM and Korean version of the Modified Barthel Index (K-MBI).[20] This was a prospective, common person linking study of 276 individuals. Methods were similar to those of the prior studies. After co-calibrating FIM and K-MBI and removing "invalid" data, exploratory factor analysis was conducted in which it was determined that three dimensions were present in the co-calibrated item pool: self-care, mobility and involuntary movement. The authors therefore ran three separate Rasch analyses for each dimension. For each factor, authors presented PCA, Rasch fit statistics, precision and reliability, and hierarchical structure of items and correlations of actual vs. converted scores. In this case, correlations were all greater than 0.91, meeting the criterion of 0.866 suggested by Dorans.[8] No sub-group analyses, distribution differences, or point differences between actual and cross-walked scores were presented and there was no discussion of the appropriateness of the crosswalk for individual vs. group data. Authors did note that as there were misfit items in two of the sub-scales (self-care and mobility), this may introduce error into the Rasch model. They also noted that due to the removal of invalid data, this is an "ideal" sample which should be validated in a "real world" sample.

2.9.7 Comparison of the Three Crosswalk Methods

The three methods of creating a crosswalk have different strengths, weaknesses and underlying concepts.

Expert panel linking has the advantage, which can also be considered a disadvantage, of using expert opinion vs. relying solely on statistical analysis. It is subject to the small sample size of an expert panel and the influence of personal bias, which will impact results. Expert input is used in the instrument validation processes to provide input on content and face validity. This can be informal or more formal

such as the methods described in a Delphi process. A more formal and step-wise process for expert panel linking process may mitigate the potential disadvantages of subjective input.

Another weakness of expert panel equilibration in the case of a FIM to SCIM III crosswalk, is that each item and score need to be compared and an appropriate "match" identified. As the number of items in FIM (11) and SCIM III (16) addressing voluntary motor function are not equivalent and SCIM III contains items that are not represented in FIM (e.g. floor and car transfers), some items may not be used. In addition, scoring on each instrument differs, with 7 FIM scores and a varying number of SCIM III scores (ranging from 2-9). Therefore, in some cases scoring will be collapsed and individual items that do not match may not be considered, losing potentially valuable data. In addition, as noted by Velozo[89], if individual items and scores are not used, this may alter (improve or decrease) instrument reliability.

Equipercentile linking uses total scores, so no information is lost in creating the crosswalk. In addition, it is a simple method and fairly easy to understand. Disadvantages are that it is based on classical test theory, relying on total scores from ordinal measures. It does not consider differences in item difficulty and is not sample and test independent. Also, in addition to inherent instrument error, additional error can be introduced in the cross linking function due to rounding.

Rasch linking linearizes measures and creates a crosswalk by assessing "fit" of actual vs. predicted scores to a unidimensional Rasch model. Thus, it is a more complex model assessing multiple aspects of the crosswalk and linking the crosswalk to a linearized measure. Disadvantages are that the model is difficult to understand without a statistical background.

Using three conceptually different linking approaches will allow comparisons between methods. If all three methods demonstrate a strong link, this is indicative of a strong overall linking relationship. Comparison of methods may highlight different strengths and weaknesses between methods or the

appropriate use in a given situation. Although other studies have compared crosswalk methods, to our knowledge comparison of the three methods outlined above has not occurred, nor has crosslinking been conducted in SCI.

CHAPTER

III. METHODS

In this section, a summary of the research design, samples used for the development and validation databases, as well as data preparation and analysis methods are presented. Crosswalk development, analysis and validation is organized by Specific Aim.

The word database will represent data from a given database (SWISS, US SCIM III reliability study - Anderson, RHSCIR). Dataset represents the entire set of data within a database including individual participant data at multiple time points (if available). A data point refers to the final, complete data included in the analysis after randomization, which represents a single exam for a participant at a single time point.

3.1 Research Design

The research design was a retrospective analysis of two existing databases (SWISS and RHSICR) to establish a FIM/SCIM III crosswalk. Common person linking, in which data for both instruments are collected on the same individuals, was used, where both FIM and SCIM III were collected within seven days of each other.

To assess a sufficient number of samples within a reasonable timeframe, retrospective data were used, but an additional strength of this design is that it reflects “real-world” use of and error within FIM and SCIM III assessments that were not part of a defined clinical study. In a prospective study design, knowledge of the intended research purpose may bias results. Common person linking is a strong design as it allows for direct comparison of actual and cross-walked scores. With retrospective data collection, order effects cannot be accounted for, but if the order is documented one could

determine if order effects are present. As these are measures of ADLs assessed by observation, when FIM and SCIM are assessed in close proximity (e.g. within the same day), an individual may be more fatigued during the assessment of the second measure. Separating the two assessments in time, would minimize this problem, but if the assessments are too far apart, differences in scores between instruments may be due to actual change.

FIM and SCIM III assessments collected within seven days of one another were used, as per Velozo et al.[89] This time frame is chosen as one in which changes between FIM and SCIM III scores are less likely than with longer time frames, but will still result in a high number of data points. Although there is a risk of fatigue when both COAs are collected by observation on the same day, these data will be included in the sample. Although exact time frames between exams are not available from the SWISS data base, per the investigator all assessments occurred within seven days, and most occurred within one to three days of one another. Days between assessments and order of exams was available for RHSCIR and Anderson data.

3.2 Population and Sample

See

Table 6 for a summary of all databases.

3.2.1 SWISS Database

The SWISS database encompasses data from four SCI rehabilitation centers in Switzerland, with comparable systems of care and rehabilitation approaches. As part of an initiative to establish a new payment system for SCI rehabilitation, the Swiss DRG AG (a joint institution of the service providers, insurers and cantons in the Swiss healthcare system) sponsored collection of FIM and SCIM III data on the same participants, predominantly with traumatic SCI, during inpatient rehabilitation. SCIM III is routinely collected at all rehabilitation centers in Switzerland and this network was established to

compare the use of FIM and SCIM III. As data were collected at calendar weeks, the time since injury was variable and was not collected in the dataset.

FIM training was via a "commercial workshop" presumably by Uniform Data Systems. FIM is not routinely collected in the SWISS SCI centers. No specific SCIM III training occurred as all SCIM raters routinely collect SCIM III data and thus had experience with the instrument.

3.2.2 Rick Hansen Spinal Cord Injury Registry

The RHSCIR database, established in 2004, is based in Canada. There are 31 clinical sites, representing 9 of 10 Canadian Provinces, and three sites outside of Canada. As of September 2018, the registry contained data on 7196 participants with data collected during acute hospital admission and up to 10 years post injury. It is estimated that this database captures 60-70% of traumatic SCI cases in Canada. The main purpose of this registry is to subsequently: 1) answer research questions related to the epidemiology of the injury and the effectiveness of current and proposed treatments; 2) evaluate the quality of care delivery and facilitate the implementation of best practices; 3) facilitate the implementation of clinical trials; and 4) create a participant database for future SCI clinical research studies.

FIM training is standardized and provided by Uniform Data Systems. SCIM III training is via webinar and/or in person training when necessary. A guidance document was also created to improve consistency (https://scireproject.com/wp-content/uploads/SCIM_Toolkit_Printable-1.pdf). Collected ISNCSCI data (severity and level of injury) was used for this project. ISNCSCI is completed by spine surgeons, physiatrists, physical therapists (PT), clinical nurse specialists or registered nurses (RN). Most data collectors are trained via the online American Spinal Injury Association (ASIA) training modules, while some sites/personnel receive additional in person training. Both FIM and SCIM III are routinely collected for the RHSCIR registry.

3.2.3 Anderson Database

The Anderson database was part of a multi-center study, to assess the reliability and validity of SCIM III in the US (see Section 2.2.1), with data collected between 2008 and 2010. As part of the assessment of validity, FIM data were also collected in order to compare responsiveness of FIM and SCIM III. FIM data are collected routinely in the US, while SCIM III data were collected for the purposes of this study.

FIM training was standardized and provided by Uniform Data Systems. No SCIM III training was provided and no guidance manuals were provided. SCIM III developers believe that no guidance is necessary beyond the descriptors provided in the instrument and as this study was intended to emulate the way instruments are typically used, no training was provided. FIM is routinely collected in US SCI sites, while SCIM III is not routinely collected.

3.2.4 Comparison of Databases

Data from all databases used in this study were collected during different years and at different time points in a subject's episode of care. Additionally, the data are from three different countries with varying lengths of stay.

Data on length of stay have not been published for each of these databases for the time periods from which the data for this study were collected. However, looking at current data on lengths of stay in each of these systems of care provides some insight. As of 2004, the average rehabilitation length of stay in centers contributing to the EMSCI data (SWISS data contributors are a part of this network) was 140 days (personal communication Armin Curt, April 2018), while the median in Canada is currently 73 days (personal communication Suzanne Humphreys, March 2019) and the median length of stay in the US based SCIMS was 34 days in 2018.[36] However, the differences in length of stay and time period over which data were collected should not impact the analyses, as relationship between FIM and SCIM III at

any time point is relevant, not changes in FIM and SCIM III in relation to systems of care or rehabilitation.

FIM is collected routinely in the US, while SCIM III is collected in the SWISS centers and both COAs are collected in RHSCIR. When a measure is collected only for research purposes, in which data collectors do not use the COA as part of routine clinical practice, reliability may be impacted. However, at least for the data collected in Anderson, reliability was strong (Pearson correlation = 0.81)[1]. Reliability in SWISS and RHSCIR data used for this study has not been directly assessed.

SWISS and Anderson data were collected by observation. RHSCIR FIM was collected by interview, while SCIM III was collected by observation, interview or the method was not indicated. Based on a 2018 publication by Itzkovich, et al.[44] (summarized in Section 2.2.2), SCIM III by interview should be used cautiously due to low interrater reliability for some items. For this analysis, only SCIM III by observation was used.

Typically, when comparing data between two samples, one would determine if there were any significant differences in the distribution and demographics of the datasets. However, no demographic data were available for the SWISS data so it was not possible to make this comparison. One potential impact is that it is not known how soon after injury the conjoint exams occurred. If, using the criterion of seven days or less between exams, these exams occurred early post injury and were seven days apart, it is more likely that differences between FIM and SCIM III could be related to changes in function over the seven days.

Although there were differences between the databases, the difference that is most likely to be impactful is the method of data collection.

Table 6: Summary of databases used in this study

	SWISS	RHSCIR	Anderson
Use in crosswalk study	Crosswalk development	Crosswalk validation	Crosswalk Validation (Methods 2 and 3)
Number of participating centers	4	30	19
Country	Switzerland	Canada	United States
Intended purpose of database	Research: FIM and SCIM III comparison	Registry	Research: SCIM III reliability study
Timespan of data collection	2017 - 2018	2014 – 2018	2008 - 2010
Collected data used for crosswalk	FIM, SCIM III	FIM, SCIM III Age, gender, profession of evaluator, days from injury to exam, neurological level of injury derived from ISNCSCI	FIM, SCIM III (two SCIM III assessments as this was a reliability study) Age, gender, days from injury to exam, neurological level of injury derived from ISNCSCI
Method of FIM/SCIM III data collection	Observation	FIM: clinician interview, SCIM III: clinician interview, observation or not indicated	Observation
Profession of FIM/SCIM III data collectors	PT, OT, RN	PT, OT, RN	PT,OT,RN, clinical research staff, physiatry
Time at which FIM/SCIM III data are collected	By calendar week (cross sectional): 2017 calendar weeks 28, 35, 42 and 49; 2018 calendar weeks 4 and 10	Rehabilitation admission and discharge	Within one week of rehabilitation admission, within one week of rehabilitation discharge
Number of subjects	663	557	390
Number of longitudinal data collection time points	1-5	1-2	1-2

ISNCSCI= International Standards for Neurological Classification of SCI, PT= Physical Therapy, OT=Occupational Therapy, RN= Registered Nurse

3.3 Clinical Outcome Assessments

3.3.1 Functional Independence Measure

FIM is a generic tool to assess burden of care, and the most widely used assessment of ADLs in rehabilitation. FIM is collected by the largest SCI data base in the US, the SCIMS database, and is collected in numerous other databases including the North American Clinical Trials Network - NACTN (FIM and SCIM II) and RHSCIR (FIM and SCIM III). Instrument details and literature review are provided in Section 2.2.1.

3.3.2 Spinal Cord Independence Measure

SCIM III is an SCI specific ADL assessment, with increasing use both clinically and for research. SCIM III is widely used in SCI registries such as EMSCI (the largest registry in Europe) as well as NACTN (SCIM II) and RHSCIR. SCIM III is used extensively in investigator sponsored SCI studies and recently has been used as a secondary endpoint in industry sponsored studies of drug and biological agents.[7] Instrument details and literature review are provided in Section 2.2.2.

3.4 Funding

This project is funded by the University of Texas Medical Branch, Center for Large Data Research and Data Sharing in Rehabilitation (sub-award # 18-84411-9), and the Rick Hansen Institute (fees associated with data access).

This project is supported in part by NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. Contents are the authors' sole responsibility and do not necessarily represent official NIH views.

3.5 Ethics Review

The Colorado Multiple Institutional Review Board determined this research to be exempt from IRB approval, under Category 4 (secondary data) on 24 June 2019, protocol # 19-1425.

3.6 Data Use and Preparation

Voluntary motor items for FIM and SCIM III were used to create crosswalks (Table 1 and Table 3). Reed et al.[66] define volitional performance as "voluntary, task-specific physical actions contributing to independence in activities of daily living" and are "repeatable movement activities", which are referred to here as voluntary motor function. Items related to bowel, bladder, respiratory function or cognition were not incorporated as they represent different constructs or domains that are influenced by involuntary motor activities due to autonomic nervous system function.[66] Use of toilet is included in the SCIM III sub-scale of respiration and sphincter management. This item assesses perineal hygiene, adjustment of clothing, and application of napkins or diapers which is a voluntary motor activity, thus this item was retained from SCIM III, as well as a similar item in FIM. In summary a total of 11 of 18 FIM and 16 of 19 SCIM III items were used.

3.6.1 Missing Data and Out of Range Values

Data entries with missing data or out of range values (determined by values outside the range of values listed above) were not used in the analysis. The percent of missing data for both databases was assessed.

3.7 Creation of Crosswalks and Data Analysis

Three theoretically different methods were used to establish FIM/SCIM III crosswalks and the best method of the three was identified, based on the criteria assessing crosswalk strength (see Specific Aim 3 below), in both the development and validation data sets. SPSS 24 was used to develop the crosswalks(s) for Methods 1 and 2, and Winsteps version 4.5.1 for Method 3. Assessment of crosswalk(s) strength was conducted in SPSS. The methods and individual steps are outlined below.

3.7.1 Specific Aim 1

Assess the number of dimensions in a combined FIM and SCIM III voluntary motor function item bank.

Hypothesis: Assessments of dimensionality will support the use of a single crosswalk for each crosswalk method.

Null hypothesis: Assessments of dimensionality will support the use of multiple crosswalks for each method.

The hypothesis will be accepted if exploratory factor and/or Rasch analysis support a single underlying construct.

The null hypothesis will be accepted if exploratory factor and Rasch analysis support multiple underlying constructs.

A core tenet of linking is that the measures to be linked must represent a single and similar construct. For example, if one measure assesses physical symptoms of depression and another measures cognitive symptoms of depression, the two instruments should not be linked. If two measures both include physical and cognitive symptom of depression as separate factors, two separate crosswalks for these dimensions can be created. In addition, unidimensionality is a requirement for Rasch analysis.

As discussed in the literature review (Chapter 2), FIM and SCIM III assess a similar underlying construct, ADLs. Limiting items of both measures to voluntary motor function further ensured similarity of content. However, due to prior publications noting multiple constructs in FIM [19,20] and item misfit in SCIM III[46,66] it is possible there are multiple domains within the voluntary motor items. Thus, the number of underlying constructs were explored in the development database.

The determination of the number of factors is along a spectrum, as some degree of multidimensionality exists. The assessment of the dimensionality is made in the context of the work and

considers the strengths of the dimensions. The factor analysis approach and considerations are presented below, while the Rasch assessment of dimensionality is presented in the Rasch section below.

Factor Analysis

EFA was conducted in SPSS to determine if FIM and SCIM III voluntary motor items assess a single or multiple underlying constructs, which would necessitate multiple crosswalks for each method.[113]

Although factor analysis options in SPSS ideally are used for continuous variables, Baglin[121] demonstrated that differences between methods for ordinal vs. continuous data, when applied to ordinal data, identified the same number of factors, although the loading was slightly different on each factor. For the purpose of this dissertation to identify the number of latent variables, the use of SPSS for ordinal data was considered to be sufficient.

The outline provided the University of California, Los Angeles statistical consulting group for factor analysis in SPSS was followed.[113] The hypothesis was tested by using principal axis factoring, with oblique rotation methods (Promax) which is appropriate when it is likely the factors are not completely independent and may be correlated. The default number of maximum rotations (25) was used. The number of initial factor extractions was based on the scree plot and the number of factors with eigenvalues with totals >1.0.

In assessing the number of latent traits we largely followed the methods of Noonan[75] and Reeve[75] (except where indicated) and considered the following:

- Scree plot: The location of the "elbow".[75,112]
- Magnitude of the eigenvalues: A minimum of 20% of the variability in the first eigenvalue supports a single underlying construct.[75,112]

- Ratio of the first and second factor: A ratio > 4 supports a single underlying construct.[75,112]
- Total variance explained: 60-80% of the total variance explained by the chosen number of factors.[113]
- Pattern of factor loadings in the pattern matrix (reported as regression coefficients when oblique rotation is used): Are items clearly loading on one factor or another or are multiple items loading on multiple factors?[75,112]
- The correlation between factors: High correlations indicate overlap between factors, where low correlations indicate distinct factors with minimal overlap.[112]

3.7.2 Specific Aim 2

Crosswalk(s) for FIM and SCIM III voluntary motor function items will be created using three conceptually different methods: expert panel linking, equipercentile linking and Rasch analysis co-calibration. Correlations between actual and cross-walked scores using the crosswalk(s) for each of the three methods will be assessed.

Hypothesis: Correlations will exceed established criteria (0.866)[8] using the crosswalk(s) for at least one of the three methods.

Null Hypothesis: Correlations will not exceed established criteria (0.866)[8] using the crosswalk(s) for any of the three methods.

3.7.2.1 Method 1: Expert Panel Linking

Step 1: Establish the crosswalk

Three experts in the field of SCI (Linda Jones, PT, MS; Gale Whiteneck, PhD; Vanessa Noonan, PT, PhD) independently created a table identifying items and scores from FIM and SCIM III with shared traits. The experts are collaborators on this project who have each participated in SCI research for a

minimum of 15 years and have multiple publications on SCI functional outcomes. Differences between the three experts were discussed and adjudicated. From this table, a common reduced scale, the Expert panel FIM/SCIM III (EFS), was created. FIM consists of 11 voluntary motor items with scores from 1-7 (7 levels) and SCIM contains 16 voluntary motor items with the maximum range of scores varying per category (from 2-9 levels) (Table 1). As the number of items and scoring levels differ between FIM and SCIM III, some items and scores were collapsed, resulting in a reduced set of items and scores. The EFS table was reviewed by experienced clinicians (PTs and OTs) who use both FIM and SCIM III in their clinical practice and based on their recommendations, two minor adjustments were incorporated.

Step 2: Recode FIM and SCIM III scores

FIM and SCIM III individual item scores collected on the same individuals from the SWISS database were recoded from the individual FIM and SCIM III items (respectively) to the EFS item scores. Total scores for EFS FIM and EFS SCIM were calculated from the new EFS FIM and EFS SCIM individual item scores. Coding can be found in Appendix C.

3.7.2.2 Method 2: Equipercentile Linking

Step 1: Rank ordering and crosswalk table

Total FIM and SCIM III voluntary motor scores (across 11 FIM and 16 SCIM items) were rank ordered separately. Scores were then lined up side by side based on percentile rank from the total voluntary motor score.

A SCIM equivalent score for FIM was created and a FIM equivalent score for SCIM was created. For example, raw total FIM scores at the 50th percentile were lined up with raw total SCIM III scores at the 50^{th} percentile and vice versa. For a given individual, SCIM III and FIM scores may not fall at the same percentile rank. Due to differences in score ranges (range of voluntary motor total scores for FIM = 11-77, SCIM = 0-65), and rounding (the exact number of FIM scores at the 50^{th} percentile is not equivalent

to the exact number of SCIM scores at the 50th percentile), two crosswalk tables were created, one for the conversion of SCIM III to equipercentile FIM (EQFIM) and one for the conversion of FIM to equipercentile SCIM (EQSCIM).

Step 2: Recode FIM and SCIM III scores

Raw total FIM scores were then recoded to EQ SCIM total scores and raw total SCIM III scores were recoded to EQ FIM. Coding can be found in Appendix C.

Step 3: Assess score distributions

Due to differences in frequency distributions for the measures, the score distributions and equipercentile relationships may appear irregular, which can create error in linking. Score distributions were visually assessed and if irregular, pre-smoothing log linear methods (smoothing the score distributions) were applied.

Step 4: Re-create equipercentile crosswalk tables

If smoothing was used, the equipercentile linking function was re-created after smoothing is applied.

3.7.2.3 Method 3: Rasch analysis using Winsteps 4.3.4

Rasch analysis (one parameter IRT) was chosen over multi-parameter models as it is a simpler model, for which model fit can be obtained and based on a prior linking study[107], resulted in an equally strong crosswalk. The steps outlined by Linacre and Velozo[89] were used. For this analysis, the group rating scale model (for polytonomous scores when like groups of items with the same scoring metric are treated on their own scale, and different groups of items with a different scoring metric are treated as different scales) was followed. This model is appropriate as FIM scoring is consistent across all items,

while SCIM III scoring differs both between some items. Where SCIM III item scoring (and descriptors) are the same, these items are grouped. The control file with coding can be found in Appendix C.

Step 1: Load and analyze FIM and SCIM III as separate Winstep files and assess Rasch output for each measure. (Run 1 and 2, Appendix C, 1.0)

Step 2: Load FIM and SCIM III in the same Winstep file, using the group rating scale model. This step places FIM and SCIM III on a common linear (logit) scale. Assess Rasch output for the common scale and measures including dimensionality. (Run 3, Appendix C, 1.0)

Item misfit and score threshold disorder were examined as part of the assessment of unidimensionality. For items that did not fit the model, removal of these items was considered using the criteria of 0.5-1.7 mean square fit as suggested by Wright and Linacre[110] for clinical observation. For score threshold disorder, re-scaling with a reduced number of scores was considered.

Step 3: Anchor separate FIM and SCIM III items analyses to item and rating scale measures from the co-calibrated analyses by creating an item anchor file (IAFILE in Winsteps) and rating scale structure file (SAFILE in Winsteps). This file will be used in Runs 4 and 5 (Appendix C, 1.0).

Step 4: Generate a Table of Measures (Table 20.1 in Winsteps) separately for FIM and SCIM III (Run 4 and 5, Appendix C, 1.0)

Each of these analyses produced a separate table, which connects total FIM and total SCIM III raw scores to person measures in logits. From this step a final crosswalk of raw FIM to raw SCIM III scores was generated, from the two Rasch analyses of FIM and SCIM III data (Step 3), anchored to the co-calibrated item and rating scale measures.

Step 5: Generate crosswalk tables

Each score in FIM and SCIM III was "matched" based on logits. Where matched logits between measures were not exact, the score to the closest logit was used. This can result in multiple scores from one measure matched to another, or cases where there is no match for a particular score. In cases where there was no direct match and logit scores are equidistant, a 0.5 was added to the best match score.

In prior steps, the person-ability measures from separate FIM and SCIM III Rasch analysis were anchored to item and rating scale measures. This allowed linkage of raw scores and person measures to create a conversion table (using natural logarithm units or logits), so that raw total FIM scores could be linked to the raw total SCIM III scores and vice versa. This is similar to equipercentile linking, except that linearized logits are used vs. raw scores and equipercentile rank.

Step 6: Recode FIM and SCIM III scores

Raw total FIM scores were recoded to R SCIM and raw total SCIM III scores were recoded to R FIM scores using the previously created crosswalk table. Coding can be found in Appendix C, 1.1.

The hypothesis was tested using a Pearson correlation (see Appendix B for formula) to compare actual and cross-walked scores and accepted if the correlation meets or exceeds 0.866 for all crosswalk(s) in a single method.

The null hypothesis was accepted if the correlation was less than 0.866 for any crosswalk(s) in a single method. Although Spearman is the more appropriate correlation for ordinal data, Pearson was used for comparisons with existing data.

Method 1: The correlations between the conversion of SCIM to EFS SCIM and FIM to EFS FIM.

Method 2: Correlations for FIM to EQ FIM and SCIM to EQ SCIM.

Method 3: Correlations for FIM to R FIM, SCIM to R SCIM.

As the goal is to provide a crosswalk for linking individual scores, the criterion of 0.866 proposed by Dorans[72] in order to achieve at least a 50% reduction in the uncertainty of predicting a score using the crosswalk was used (see Appendix B for formula).

Additional criteria (below) were assessed for methods with a correlation exceeding 0.866. Although the Pearson correlation coefficient was the primary criterion, the criteria below were also assessed if correlation coefficients were similar to support the use of one crosswalk method over another in a given situation or population. Ideally raw scores and cross-walked scores (or in the case of Method 1 EFS SCIM to EFS SCIM) were similar. The assessments below were intended to provide additional information about what and where differences lie.

- Score distributions between actual vs. cross-walked scores: means, standard deviations, skewness, kurtosis were compared.
- Distribution of differences between actual and cross-walked scores: Bland-Altman plots were generated and visually inspected.[83]
- Amount of difference between the two distributions: Cohen's effect size (see Appendix B for formula) using the criteria of 0.8 = large, .5 = medium, and .2 =small.
- Point differences between actual and cross-walked scores. A threshold of 75% was used for scores that differ between actual and converted scores for each COA, by ½ a standard deviation.

As discussed in Chapter 2, a variety of methods have been used to examine point differences between actual and cross-walked scores. Although using MCID as a "cut-off" for differences in actual versus cross-walked is appealing, the MCID for voluntary motor scores is not known, ½ of a standard

deviation was used as the criteria for point differences. The cut off used by others of 75% of these data within ½ a standard deviation was used as this is ¾ of the individual conversions.

It was not possible to assess sub-group invariance in the development database as no information on demographics are available.

Correlations between cross-walked FIM and SCIM III score were obtained for all three methods, to ensure correlations between the instruments were maintained.

3.7.3 Specific Aim 3

Validate the three crosswalk methods in a separate dataset.

Hypothesis: Correlations in the validation dataset will exceed established criteria (0.866) using crosswalk(s) for each of the three methods.

Null Hypothesis: Correlations in the validation dataset will not exceed established criterion (0.866) using crosswalk(s), for each of the three methods.

The hypothesis was tested using a Pearson correlation to compare actual and cross-walked scores in the validation dataset and accepted if the correlation met or exceeded 0.866 for all crosswalk(s) in a single method.

The null hypothesis was accepted if the correlation was less than 0.866 for any crosswalk(s) in a single method.

Additional criteria assessed for Aim 3 were replicated in the validation dataset, for methods with a correlation exceeding 0.866.

In addition, sub-group invariance using standardized mean difference (SMD – see Appendix B for formula) was assessed across gender, broad age groups, and level of injury (tetraplegia defined as a

single neurological level from C1-C8, paraplegia single neurological level between T1 – S5). Cut off values greater than 0.11[75] and 0.8[81] are commonly used in crosswalk literature to indicate that sub-group invariance is present. As the intent is to use the crosswalk for both individual and group analysis, the stricter criteria of values exceeding 0.08 were used.

Based on the primary criterion of the correlation between the actual and cross-walked scores and the secondary criteria outlined above in the development and validation datasets, the crosswalk method that best meets these criteria was identified. The hypothesis was that the Rasch model would produce the crosswalk with the highest correlation as it is a modern psychometric technique, based on the latent trait model and linearization of measures.

3.8 Invalid Data

It is possible that some collected data may indicate unexpected behavior of a given individual. For example, if an individual scores 0 “unable to ascend or descend stairs “on SCIM III, but scores a 7 “complete independence” on FIM, these data would be considered unexpected.

Velozo et al.[89], Li et al.[122], and Hong et al.[123] removed what was considered to be invalid data (removing 7%, 26%, and 8% respectively) prior to creating a crosswalk using Rasch analysis.

Data used in the FIM/SCIM III crosswalk reflect “real world” data and errors in data collection and documentation, replicating how the crosswalk will be used in the future. Thus, simply eliminating cases with unexpected behavior was carefully considered. The percentage and pattern of invalid data in the development database (SWISS) was examined using two different approaches, to determine if invalid cases should be eliminated for all three crosswalk methods. If the percentage of invalid cases was less than 10%, these cases were retained in the data set. Ten percent was chosen as it is unlikely that retaining up to 10% of cases with invalid data will significantly impact the crosswalk.

The first method was based on scores differences between FIM and SCIM III exceeding three standard deviations, which were considered invalid. The second method was based person fit on Rasch output in Winsteps. Persons with infit or outfit ≥ 3.0 were identified, based on criteria from Malec et al.[111]

In summary, the underlying construct's both within and between FIM and SCIM III were assessed. A single or multiple crosswalks (depending on the identified construct(s)), for three theoretically different approaches were developed. After evaluating and validating the crosswalk(s) for all three methods the optimal method was identified.

CHAPTER

IV. RESULTS

Results are presented by summarizing the data and demographics used in the analysis followed by the assessments of dimensionality (Aim 1). Crosswalk development, validation and assessment (Aims 2 and 3) are presented in aggregate in the following order: crosswalk development by method, primary crosswalk assessment (correlation), secondary crosswalk assessments and additional analysis, followed by comparisons of methods, datasets and outcome measures.

Dimensionality was assessed and the crosswalks developed using 662 data points from the SWISS dataset. Crosswalks were then validated in both the Anderson (n=119) and RHSCIR (n=133) datasets.

4.1 Final Dataset

Final sample sizes, demographics and characteristics of all datasets are in Table 7.

4.1.1 SWISS

The original dataset contained 985 FIM/SCIM III data points, from 663 unique participants, who were assessed from 1-5 times over the course of their rehabilitation stay (Figure 6), at cross sectional time points, collected approximately every 6-7 weeks. Data from two data points contained a missing or out of range value, leaving a complete data set of 983/985 data points from 662/663 unique participants. Out of range values were identified based on the range of values for each assessment and item (Table 1). Approximately 1% of data points exceeded three standard deviations from the mean, and 6% demonstrated person misfit in Rasch analysis. As both analyses showed less than 10% item misfit, all data were retained for analysis.

Where a single complete data set per participant existed, this data point was used in the analysis. Where multiple complete data points existed per participant, the data points were randomized so that only one complete, independent data point was used per subject. No demographic data were available. As the assessment periods were cross sectional, it was not possible to determine the time point in rehabilitation at which the data were collected. However, one can estimate that at the first cross sectional data collection time point in which data for an individual is captured, the average time from admission was between 0-7 weeks or approximately 3.5 weeks (25 days), with a second time frame equating to 10 weeks, (70 days) and a third at 16.5 weeks (116), etc.

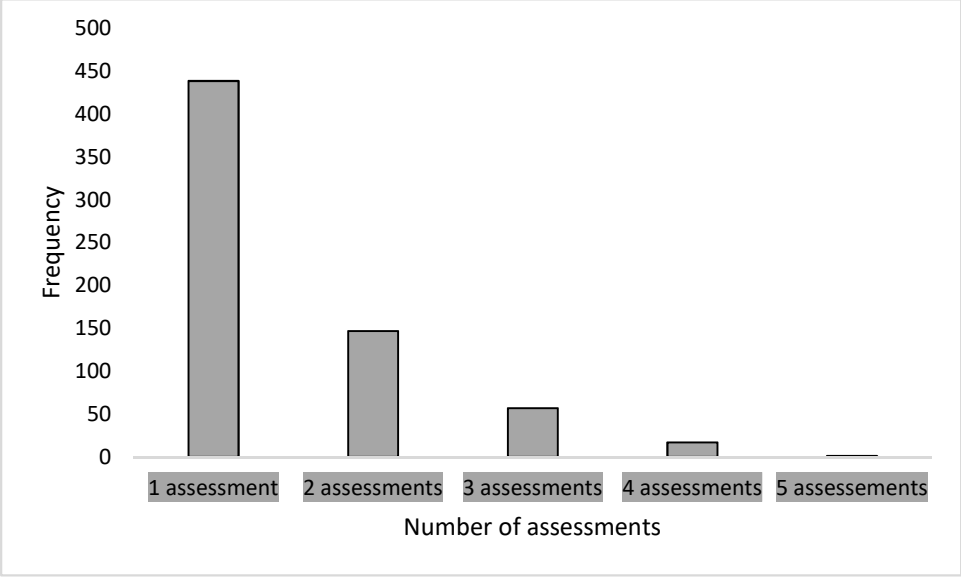


Figure 6: SWISS Frequency of number of assessments

4.1.2 Anderson database

This database originally contained data on 390 unique participants, at rehabilitation admission and discharge, resulting in 780 data points. Each subject had two SCIM III assessments by different assessors as this was a SCIM III inter-rater reliability study. After eliminating missing or out of range

values, data for which it was not possible to calculate the time frame between FIM and SCIM III, and data for which the FIM locomotion mode was not indicated, a dataset of 119/390 (31%) data points remained. Where only a single complete data point existed for a SCIM III assessor (assessor one or two), or admission vs. discharge, this data point was used in the analysis. Complete data points were randomized by SCIM III assessor, then by admission vs. discharge. Randomization was weighted to approximate alignment with the SWISS dataset. The first two cross sectional time points in the SWISS (between 0-7 weeks or approximately 3.5 weeks [25 days], with a second time frame equating to 10 weeks, [70 days]) constitute the majority of the SWISS data. Assuming an approximate length of stay in Anderson of 34 days based on the length of stay in Model Systems[36], 80% of 25 days (first cross section in SWISS) = 20 days, 20% of 70 days = 14 days, 20+34= 34 days. Thus, the intended target was 20% admission and 80% discharge data. Randomization resulted in 19% (23/119) admission and 81% (96/119) discharge data points. An additional notation is that in some cases SCIM III exams occurred across multiple days. Based on available data, this occurred in four cases randomized to admission (range 1-3 days), and eight cases randomized to discharge (range 1-4 days).

4.1.3 Rick Hansen Spinal Cord Injury Registry

The original dataset contained 557 participants with rehabilitation admission and discharge data for a total of 1,114 data sets. A total of 133/557 (24%) data points remained after eliminating data points for the following reasons: missing or out of range values, unable to calculate time frame between FIM and SCIM III, SCIM III collected by questionnaire (vs. observation). After randomizing to ensure independent data, 75/133 (56%) admission data points and 58/133 (44%) discharge data points were used in the analysis. As the length of stay in Canada was approximately 73 days (close to the second data collection point in SWISS), data were randomized with an approximate 50/50 split admission to discharge.

4.1.4 Comparison of databases

Table 7 summarizes and compares demographics and key data from each dataset, when available. Unless otherwise stated, comparisons are between Anderson and RHSCIR. Data for SWISS are only available for FIM scores and mode and SCIM III scores for the time point used in the analysis.

- Demographics: There were no significant differences in gender or age.
- Injury Characteristics: There was a significant difference in the level of injury for admit and discharge combined (all time points in table) as well as for rehabilitation discharge alone. The RHSCIR dataset included a significantly higher percentage of individuals with tetraplegia (66%- all time points, 69% at discharge) vs. Anderson (47% - all time points, 57% at discharge). Severity of injury was not significantly different between datasets, when analyzed for motor complete vs. incomplete and ASIA Impairment Scale A (motor and sensory complete), B (sensory complete), C (motor incomplete), D (motor incomplete with more motor function than C). (See the American Spinal Injury Association website for complete descriptions)[124].
- Data Characteristics:
    - *Randomization:* There was a significant difference in the percentage of data points randomized to admission (Anderson 19% vs. RHSCIR 56%) vs. discharge (Anderson 81% vs. RHSCIR 44%) which was intentional as noted above.
    - *Length of stay:* Differences in length of stay (LOS) were significant for rehabilitation length of stay (Anderson 39.40 ± 24.87 vs. RHSCIR 97.09 ± 54.81) and total (acute + rehabilitation) length of stay (Anderson 68.86 ± 49.61 vs. RHSCIR 138.95 ± 75.91) with an average difference of 58 days for rehabilitation LOS and 70 for total LOS.
    - *Days between injury and exam:* date of injury to discharge exam for both FIM (Anderson=68.4 ± 54.43, RHSCIR= 146.76 ± 82.20) and SCIM III (Anderson 66.54 ± 55.11,

RHSCIR 147.17 ± 82.16) were significantly different which is consistent with the longer length of stay noted above.

- *Days between FIM and SCIM III exams:* Significant differences were noted only at discharge (Anderson 1.90 ± 2.36, RHSCIR .41± 2.70).
- *FIM and SCIM III characteristics:* Significant differences were found for FIM mode (walk vs. wheelchair) between all three datasets, with SWISS having a lower percentage of walk (9%) relative to Anderson (27%) and RHSCIR (31%). SCIM III total scores at discharge were significantly lower in Anderson (28.25 ±12.28) than RHSCIR (32.10 ±10.06).

Table 7: Demographics, injury and data characteristics

	SWISS	Anderson	Rick Hansen Registry	*p*
Sample size (n)	662	119	133	NA
DEMOGRAPHICS				
Gender [percent, n]	Not available	Male = 75% (89/119), Female= 25% (30/119)	Male = 76.7% (102/133), female =23.3% (31/133)	0.73
Mean age (years) (mean ± SD)	Unknown	45 ±=17.54	47.09 ± 18.03	0.37
INJURY CHARACTERISTICS				
Level of Injury (percent, n)				
Tetraplegia vs. paraplegia **ADMISSION**	Not available	Tetraplegia= 52% (12/23), Paraplegia 48% (11/23)	Tetraplegia= 64% (42/66), Paraplegia=36% (24/66)	0.33
Tetraplegia vs. paraplegia **DISCHARGE**	Not available	Tetraplegia =43% (15/35), Paraplegia = 57% (20/35)*	Tetraplegia= 69% (25/36), Paraplegia= 31%(11/36)*	**0.02**
Tetraplegia vs. paraplegia **ALL TIME POINTS**	Not available	Tetraplegia= 47% (27/58), Paraplegia=53% (31/58)*	Tetraplegia = 66% (67/102), Paraplegia= 34% (35/102)*	**0.02**
Severity (percent, n)				
Motor complete vs. motor incomplete **ADMISSION**	Not available	Motor complete = 52% (12/23), Motor incomplete = 48% (11/23)	Motor complete = 53% (36/68), Motor incomplete = 47% (32/68)	0.95
Motor complete vs. motor incomplete **DISCHARGE**	Not available	Motor complete = 38% (18/47) Motor incomplete = 62% (29/47)	Motor complete = 51% (53/104) Motor incomplete = 49% (51/104)	0.14
Motor complete vs. motor incomplete **ALL TIME POINTS**	Not available	Motor complete = 43% (30/70), Motor incomplete = 57% (40/70)	Motor complete = 51% (53/104), Motor incomplete = 49% (51/104)	0.29
AIS A, B, C, D **ADMISSION**	Not available	A= 39%, (9/23), B=13% (3/23), C=9%, (2/23), D=39% (9/23)	A=38%(26/68), B=15% (10/68), C=12% (8/68), D=35% (24/68)	0.55
AIS A, B, C, D **DISCHARGE**	Not available	A=23% (11/47), B=15% (7/47), C=13% (6/47), D=49% (23/47)	A=36%(37/104), B=15% (16/104), C=10% (10/104), D=39%(41/104)	0.47
AIS A, B, C, D **ALL TIME POINTS**	Not available	A=29% (20/70), B=14% (10/70), C=11% (8/70), D=46% (32/70)	A=36% (37/104), B=15%(16/104), C=9.6% (10/104), D=39% (41/104)	0.69

Table 7 con't

	SWISS	Anderson	Rick Hansen Registry	*p*
DATA CHARACTERISTICS				
Randomization (percent, n)				
Admission vs. discharge	NA	Admission = 19% (23/119) Discharge= 81% (96/119)*	Admission =56% (75/133) Discharge= 44% (58/133)*	**<.00001**
Length of stay (days) (mean ± SD, n)				
Acute		30.09 ±42.06, (44/119)	41.96 ± 35.87 (128/133)	0.072
Rehabilitation		39.40 ± 24.87(43/119)*	97.09 ± 54.81(133/133)*	**<.00001**
Total		68.86 ± 49.61(44/119)*	138.95 ± 75.91(129/133)*	**<.00001**
Days between injury (date of injury-DOI) and exam (mean ± SD, n)				
DOI to FIM **ADMISSION**		29.09 ±53.35(22/23)	44.72 ±38.76(75/75)	0.13
DOI to FIM **DISCHARGE**		68.4 ± 54.43(95/96)*	146.76 ± 82.20(58/58)*	**<.00001**
DOI to SCIM **ADMISSION**		32.68 ± 53.71(22/23)	46.75 ± 38.95(75/75)	0.18
DOI to SCIM **DISCHARGE**		66.54 ± 55.11(95/96)*	147.17** ± 82.16(58/58)*	**<.00001**
Days between FIM and SCIM III exams (mean ± SD, n)				
FIM and SCIM **ADMISSION**	Not available	Mean = 2 ± 3.22(23/23) 6% on same day	Mean=2.03 ± 2.55(75/75) 1.5 % on same day	0.96
FIM and SCIM **DISCHARGE**	Not available	Mean= 1.90 ±2.36 (96/96)* 17% on same day	Mean= .41± 2.7(58/58)* 5.3% on same day	0.0004
FIM and SCIM **All TIME POINTS**	Not available	Mean = 0.85 ± 3.13(119/119) 14% on same day	Mean=-1.33± 2.7(133/133) 6.8% on same day	0.19
FIM and SCIM III characteristics (raw scores) (mean ± SD, n)				
FIM **ADMISSION**	Not available	25.26 ± 10.88(23/23)	29.43 ± SD= 16.65(75/75)	0.26
FIM mode	Not available	walk=35% (8/23) w/c= 65% (15/23) both = 0% (0/23)	walk= 28% 21/75) w/c = 71%(53/75) both= 1% (1/75)	0.70
FIM **DISCHARGE**	Not available	47.21 ± 13.85(96/96)	50.38 ± 20.47(58/58)	0.25

Table 7 con't

	SWISS	Anderson	Rick Hansen Registry	*p*
DATA CHARACTERISTICS				
FIM mode	Not available	walk= 25% (24/96) w/c = 69% (67/96) both= 5% (5/96)	walk=35% (20/58) w/c = 64% (37/58) both= 2% (1/58)	0.34
FIM **ALL TIMEPOINTS**	40.25 ± 19.54 (662/662)	43.00 ± 15.88 (119/119)	38.56 ± 21.09 (133/133)	0.13
FIM mode	walk= 9% (61/662) w/c= 82% (545/662) both= 9% (56/662)*	walk =27% (32/119) w/c= 69% (82/119) both = 4% (5/119)*	walk= 31% (41/133) w/c= 68% (90/133) both=2% (2/133)*	**<.00001**
SCIM **ADMISSION**	Not available	13.17 ±11.55 (23/23)	14.63 ± 13.84(75/75)	0.65
SCIM **DISCHARGE**	Not available	28.25 ±12.28,(96/96)*	32.10 ±10.06(58/58)*	**0.05**
SCIM **ALL TIME POINTS**	23.50 ± 17.21(662/662)	25.34 ± 13.49(119/119)	22.25 ± 18.44(133/133)	0.13

AIS = American Spinal Injury Association impairment scale of A (motor and sensory complete), B (sensory complete), C (motor incomplete), D (motor incomplete with more motor function than AIS C), DOI=date of injury, w/c= wheelchair

*indicates a significant difference

**This value is higher than the mean total LOS as this value is only for data points where DC SCIM/FIM were used. In some cases, LOS was not available resulting in differences. For example, in one case data on total LOS were not available but SCIM was conducted 505 days after injury, likely skewing this figure.

4.2 Assessment of dimensionality

Dimensionality was assessed in the SWISS (development) database, from a sample size of 662 data points. As discussed previously (Section 3.7.1), unidimensionality means that a test measures a single underlying trait or construct, such that differences in responses reflect item difficulty and person ability, vs. a different skill or trait.

4.2.1 Exploratory Factor Analysis

Exploratory factor analysis was conducted using two and three factors based on the location of the "elbow" in the scree plot (Figure 7). Results for a two factor solution are presented for most analyses, as the two and three factor analysis are similar with the exception of the pattern matrix which is presented for both analyses.

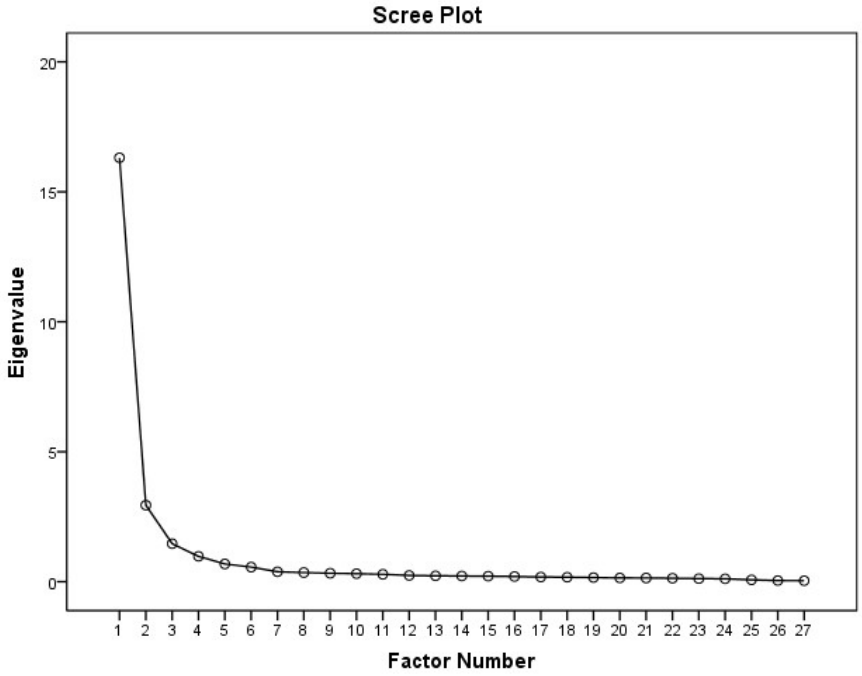


Figure 7: Scree plot

- Magnitude of the eigenvalues (Figure 8): The eigenvalue of the first factor is 60.41%, exceeding the suggested threshold of a minimum of 20%.[75,112]
- Ratio of the percent of the variance explained by the first and second factor (Figure 8): The ratio between the first and second factors is 5.53 exceeding the suggested threshold of a ratio > 4.[75,112]
- Total variance explained by the first factor (Figure 8): Approximately 60% of the variance is explained by the first factor, (Figure 8) approximating the suggested criteria of 60-80% of the total variance explained by the chosen number of factors[113], and meeting criteria of 50-60% for social science research.[113] The second factor explains 10.91% of the variance in the two factor solution. Findings are similar for the three factor solution (first factor- 59.51%, second factor – 10.04%, third factor- 4.65%).
- Pattern of factor loadings in the pattern matrix (Figure 9) (regression coefficients): In looking at regression coefficients greater than 0.3 for the two factor solution, some items clearly load on factor 1 which appears to focus predominantly on lower extremity function (FIM and SCIM stairs, SCIM transfer toilet, SCIM mobility items, transfer car and ground). Other items clearly load on factor two which focuses on upper extremity function (FIM eat and SCIM feed, FIM and SCIM groom, FIM and SCIM dress upper body). FIM walk- w/c also loads on factor 2, but not strongly (0.39). This is likely due to the fact that FIM walk-w/c encompasses a primary mode of locomotion as w/c (upper extremity function) or walk (primary lower extremity function), which cannot be separated without the FIM modifier for walk vs. w/c. While the items noted above clearly load on factor 1 (lower extremity) or factor 2 (upper extremity), a large number of other items do not clearly load on either factor (FIM bath, SCIM bath lower body, FIM and SCIM dress lower body, FIM and SCIM toilet, FIM and SCIM transfer bed, FIM transfer toilet and tub, SCIM

bed mobility). In comparing factors in three factor analysis, factor 1 appears to be a general upper and lower extremity factor with a large number of items, factor 2 is “strong lower extremity (e.g., walking and stairs) factor 3 is “strong upper extremity” (e.g. eating, grooming, dressing upper body).

Total Variance Explained			
Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	16.311	60.411	60.411
2	2.945	10.906	71.316
3	1.465	5.425	76.741
4	0.975	3.609	80.35
5	0.682	2.527	82.877
6	0.561	2.079	84.956
7	0.383	1.417	86.373
8	0.351	1.3	87.672
9	0.325	1.205	88.877
10	0.309	1.145	90.022
11	0.284	1.052	91.075
12	0.242	0.897	91.972
13	0.226	0.837	92.809
14	0.219	0.809	93.618
15	0.212	0.784	94.402
16	0.201	0.743	95.145
17	0.179	0.665	95.81
18	0.172	0.636	96.445
19	0.161	0.595	97.04
20	0.144	0.532	97.572
21	0.141	0.523	98.095
22	0.133	0.491	98.586
23	0.117	0.434	99.02
24	0.111	0.412	99.432
25	0.072	0.266	99.698
26	0.043	0.16	99.858
27	0.038	0.142	100
Extraction Method: Principal Axis Factoring.			

Factors = the number of factors extracted which is 11 (FIM) + 16 (SCIM III) = 27, Total= eigenvalues, % of Variance = the percent of total variance accounted for by each factor, Cumulative % = the cumulative percentage of variance accounted for by the current and all preceding factors

Figure 8: Exploratory factor analysis: total variance explained

Approximately 60% of the variance is explained by the first factor, with the second factor accounting for approximately 11%.

Pattern Matrix[a]		
	Factor	
	1	2
FIM Eat		1.005
FIM Groom		0.946
FIM Bath	0.410	0.537
FIM Dress UB		0.852
FIM Dress LB	0.524	0.387
FIM Toilet	0.534	0.376
FIM Xfer Bed	0.441	0.532
FIM Xfer Toilet	0.534	0.426
FIM Xfer Tub	0.542	0.398
FIM Walk-w/c		0.391
FIM Stairs	0.883	
SCIM Feed		1.001
SCIM Bath UB		0.824
SCIM Bath LB	0.498	0.412
SCIM Dress UB		0.840
SCIM Dress LB	0.493	0.406
SCIM Groom		0.976
SCIM Toilet	0.512	0.385
SCIM Mob Bed	0.358	0.578
SCIM Xfer Bed	0.506	0.390
SCIM Xfer Toilet	0.611	
SCIM Mob Ind	0.941	
SCIM Mob Dist	0.950	
SCIM Out	0.944	
SCIM Stair	1.043	-0.311
SCIM Xfer Car	0.742	
SCIM Xfer Ground	0.873	
Extraction Method: Principal Axis		
a. Rotation converged in 3		

Pattern Matrix[a]			
	Factor		
	1	2	3
FIM Eat			0.957
FIM Groom			0.786
FIM Bath	0.771		
FIM Dress UB	0.489		0.525
FIM Dress LB	0.950		
FIM Toilet	0.923		
FIM Xfer Bed	0.889		
FIM Xfer Toilet	0.944		
FIM Xfer Tub	0.955		
FIM Walk-w/c	0.313		
FIM Stairs		0.791	
SCIM Feed			1.017
SCIM Bath UB			0.690
SCIM Bath LB	0.801		
SCIM Dress UB	0.368		0.581
SCIM Dress LB	0.833		
SCIM Groom			0.933
SCIM Toilet	0.799		
SCIM Mob Bed	0.733		
SCIM Xfer Bed	0.650		
SCIM Xfer Toilet	0.556		
SCIM Mob Ind		0.879	
SCIM Mob Dist		0.903	
SCIM Out		0.891	
SCIM Stair		0.933	
SCIM Xfer Car	0.521	0.395	
SCIM Xfer Ground		0.712	
Extraction Method: Principal Axis Factoring.			
a. Rotation converged in 6 iterations.			

Factors= the number of factors extracted.Values below 0.3 are not displayed.

Figure 9: Exploratory factor analysis:pattern matrix

Regression coefficients of factor loadings for all FIM and SCIM III items, for a two and three factor analysis.

- The correlation between factors: The correlation between factors in the two factor solution is moderate-strong (0.68), For the three factor solution, correlations are moderate – strong (factor 1 and 2 = 0.69, factor 1 and 3 = 0.68 and moderate (factor 2 and 3 =0.42). These moderate - strong and moderate correlations for either the two or three factor solution indicate overlap between factors.

Although there are multiple factors that can be identified in the combined FIM/SCIM III data, the exploratory factor analysis supports the use of a single factor. Approximately 60% of the variance is explained by a single factor and a large number of individual items do not distinctly identify with one factor indicating that FIM and SCIM III share a unidimensional construct.

4.2.2 Rasch Analysis of dimensionality

As part of the Rasch analysis, dimensionality was assessed using principal components analysis as well as identifying items with misfit. Rasch PCA looks at the residuals after the Rasch dimension is extracted, to look for patterns in the residuals. There are several considerations in examining the Rasch output.

In assessing the strength of the dimensions (Table 8), the Rasch dimension explains approximately 80% of the variance, with about 39% explained by persons and 41% explained by items, which are high percentages. The second largest dimension is small and explains only 3.5% of the variance, thus the Rasch dimension explains approximately 24 times the variance than the first contrast. The eigenvalue for the first contrast is 4.60 (Table 8), which indicates that 4-5 items (out of 27) do not fit the Rasch construct. This could be due to random noise or other constructs/dimensions.

The standardized residual contrast plot for the first contrast (Figure 10) shows contrasts between items with different residual patterns and identifies clusters (see Figure 11) for item descriptors). Cluster 1 reflects items requiring lower extremity function as does cluster 2, while items in

cluster 3 reflect upper extremity function. There is no clear functional delineation between clusters 2 and 3. However, the strength of the relationship between the three clusters is quite high (Figure 12- Pearson correlation with extreme values included 0.65-0.81, dis-attenuated accounting for error variance = 0.81 - 0.96) indicating large overlap and little distinction between the clusters. If there were distinct dimensions, correlations should be less than 0.3 indicating little overlap between dimensions. So, it appears that the clusters measure the same dimension but measure it differently.

In examining item misfit using the criteria of 0.5-1.7 [110], only FIM walk/wheelchair exceeded both infit and outfit (2.35 infit, 3.23 outfit - Figure 11). This is likely due to the fact that the FIM walk/wheelchair modifier was not incorporated into the model. Thus, one can have an individual with a low score who is walking, with higher scores in other areas related to lower extremity function than if the low score was due to low functioning wheelchair use. Item misfit can be indicative of items that do not fit the unidimensional construct. With only one item out of the suggested range and given a plausible rationale, there is no strong reason to assume multidimensionality based on item misfit. SCIM bed mobility had an outfit value exceeding 1.7 (2.20), which represents outlier sensitivity (e.g. sensitive to responses in items with difficulty far from a person).

Although not related to dimensionality directly, but indicative of strengths and weaknesses of an outcome measures are mean square fit infit/outfit less than 0.5, which indicates redundant items or items that do not add information the construct being assessed. No items demonstrated mean square fit values less than 0.5 (Figure 11). Also related to utility of an outcome measure is an examination of score threshold disorder. In other words, does a higher score reflect a higher level of function and do all scores identify distinct, non-overlapping function? Three items demonstrated disordered scoring: SCIM mobility indoors, moderate distances and outdoors with disorder in scores 6 (walks with one cane) and 8 (walks without walking aids). This score disorder is likely related to the fact that scores 4-7, which

indicate the assistive device or orthosis used, can reflect personal choice this may not be always directly relate to a person's ability. Due to the use of the group rating scale, in which all FIM item scores were treated the same, disordered scoring in FIM items or SCIM items with the same scoring would not be detected.

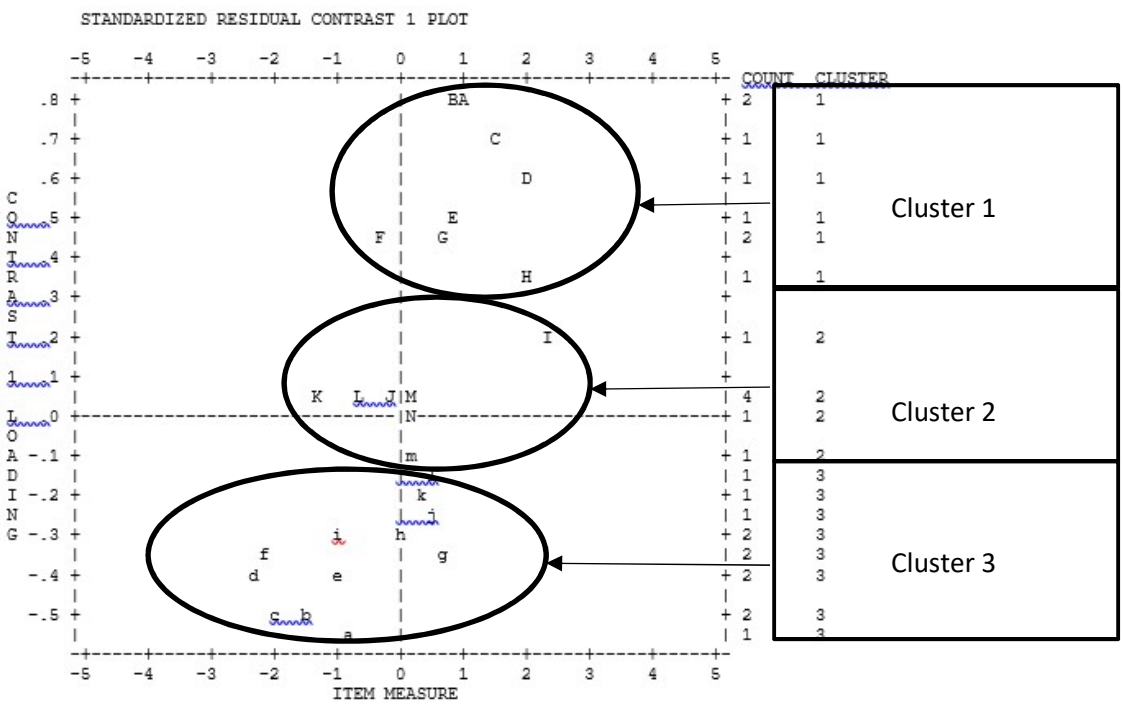
In summary, while there is some evidence suggestive of multiple dimensions in Rasch analysis of the residuals, the dimensions are not large enough to be considered distinct.

Table 8: Rasch principal components analysis: standardized residual variance in eigenvalue units

Eigenvalue for each descriptor with observed percentage of unexplained variance and expected, (expected variance if the data perfectly fit the Rasch model).

Descriptor	Eigenvalue	Observed	Expected
Total raw variance in observations	132.52	100.0%	100.0%
Raw variance explained by measures	105.52	79.6	81.2%
Raw variance explained by persons	51.02	38.5	39.3%
Raw variance explained by items	54.49	41.1	41.9%
Raw unexplained variance (total)	27.00	20.4	18.8%
Unexplained variance in 1st contrast	4.59	3.5%	17.0%
Unexplained variance in 2nd contrast	2.91	2.2%	10.8%
Unexplained variance in 3rd contrast	2.02	1.5%	7.5%
Unexplained variance in 4th contrast	1.70	1.2%	6.1%
Unexplained variance in 5th contrast	1.52	1.1%	5.6%

Extracted from Table 23.0 in Winsteps



Extracted from Table 23.1 in Winteps

Figure 10: Rasch standardized residual plot for 1st contrast

This plots shows the unstandardized "raw" loading (Contrast Loading) of the first contrast of the residuals (after the Rasch dimension is removed) of each FIM and SCIM III item against the item calibration (Item Measure). The horizontal axis is the Rasch dimension which is extracted from the data prior to the analysis of residuals. Letters "A,B,C,..." and "a,b,c,..." represents individual items (descriptors are in Figure 11 – below). The Clusters represent items clusters based on Figure 11-below. The purpose of the plot is to help determine whether the 3 clusters of items are truly measuring the same or different things.

CON TRA	CL US	LOADING	INFIT MEASURE	OUTFIT MNSQ	MNSQ	ENTRY NUMBER	ITEM	G R	
1	1	.79	.90	1.08	1.18	A 22	SModDist	F	Cluster 1
1	1	.78	.78	1.09	1.18	B 21	SMobInd	F	
1	1	.70	1.46	1.28	1.20	C 23	SOut	F	
1	1	.59	2.04	.86	.61	D 24	SStair	B	
1	1	.48	.92	.76	.74	E 26	SXfrCar	E	
1	1	.47	-.28	.75	.72	F 19	SXfrBed	E	
1	1	.47	.66	.67	.66	G 20	SXfrToil	E	
1	1	.36	1.96	.89	.39	H 27	SXfrGrnd	H	
1	2	.20	2.28	1.18	.59	I 11	FStairs	A	Cluster 2
1	2	.06	-.24	.64	.63	J 7	FXfrBed	A	
1	2	.06	-1.26	2.35	3.23	K 10	FWalkwc	A	
1	2	.04	-.60	1.09	2.20	L 25	SMobBed	G	
1	2	.03	.12	.81	.69	M 8	FXfrToil	A	
1	3	-.53	-.87	.94	1.07	a 4	FDressUB	A	Cluster 3
1	3	-.51	-1.57	1.24	1.68	b 2	FGroom	A	
1	3	-.50	-2.05	1.14	1.52	c 1	FEat	A	
1	3	-.42	-2.31	.74	.83	d 12	SFeed	B	
1	3	-.39	-.92	1.07	1.53	e 15	SDressUB	C	
1	3	-.35	-2.24	1.07	.98	f 17	SGroom	B	
1	3	-.33	.59	1.13	1.00	g 5	FDressLB	A	
1	3	-.31	.03	.98	1.44	h 3	FBath	A	
1	3	-.28	-1.06	.74	.87	i 13	SBathUB	B	
1	3	-.25	.44	1.03	.98	j 16	SDressLB	C	
1	3	-.22	.30	1.30	1.01	k 6	FToilet	A	
1	3	-.16	.48	.73	.70	l 14	SBathLB	B	
1	2	-.12	.22	1.14	1.52	m 18	SToilet	D	Cluster 2
1	2	-.01	.24	.92	.87	N 9	FXfrTub	A	

Extracted from Table 23.3 in Winsteps

Figure 11: Rasch standardized residual loadings by item for 1st contrast

Clusters from Figure 10 and associated items are identified under "entry number" and "Item" on the right hand side of the figure. The loading of each item to a cluster is under "loading". Item misfit (infit and outfit) are also displayed.

Approximate relationships between the PERSON measures

PCA Contrast	ITEM Clusters	Pearson Correlation	Disattenuated Correlation	Pearson+Extr Correlation	Disattenuated+Extr Correlation
1	1 - 3	0.6532	0.7608	0.6923	0.8133
1	1 - 2	0.8100	0.9764	0.8289	1.0000
1	2 - 3	0.8112	0.9251	0.8315	0.9609

Extracted from Table 23.2 in Winsteps
Disattenuated correlations account for measurement error, + EXTR indicates extreme values were included

Figure 12: Rasch correlations between clusters for the 1st contrast

High correlations between item clusters indicates overlap of these items, meaning they do not distinctly measure different dimensions.

As discussed in the literature review, factor and Rasch analysis take a different approach to assess dimensionality. In exploratory factor analysis, commonalities are maximized to optimize the structure of the factors, and factor loadings are correlations to the underlying latent factor(s). In Rasch analysis, the Rasch dimension is extracted, and the contrast of the residual values to the Rasch dimension are examined. The residuals are assessed to determine if they constitute a different dimension and the strength of that dimension. Thus, direct comparisons of individual output are not valid, however some patterns emerged in both analyses. In both analyses, some items are more clearly related to a factor (exploratory factor analysis) and cluster (Rasch) related to lower extremity (SCIM mobility items, SCIM and FIM stairs, SCIM and FIM toilet transfer, SCIM car and ground transfers) and upper extremity (SCIM and FIM feed/eat, SCIM and FIM groom, SCIM and FIM dress upper body, SCIM bath upper body).

4.2.3 Summary of dimensionality assessment

Specific Aim 1: Assess the number of dimensions in a combined FIM and SCIM III voluntary motor function item bank.

Hypothesis: Assessments of dimensionality will support the use of a single crosswalk for each crosswalk method.

Null hypothesis: Assessments of dimensionality will support the use of multiple crosswalks for each method.

The hypothesis will be accepted if exploratory factor and/or Rasch analysis support a single underlying construct.

The null hypothesis will be accepted if exploratory factor and Rasch analysis support multiple underlying constructs.

In summary, although both exploratory factor and Rasch analysis have components which may suggest multiple dimensions, there is more evidence to suggest a unidimensional construct. Thus the hypothesis for Aim 1 (exploratory factor and/or Rasch analysis support a single underlying construct) was be accepted and the null hypthesis (exploratory factor and Rasch analysis support multiple underlying constructs), rejected.

4.3 Crosswalk development and validation

The crosswalks were developed using 662 data points from the SWISS dataset and validated in both the Anderson (n=119) and RHSCIR (n=133) data sets. Development and validation findings (Aims 2 and 3) are summarized jointly. For all analyses, the primary analysis of interest is a comparison between FIM EFS and SCIM EFS, FIM and SCIM III raw scores and their respective equilibrated and Rasch crosswalk scores.

4.3.1 Method 1: Expert panel crosswalk

Table 9 presents the crosswalk table based on the expert panel (EFS) linking method – (detailed crosswalk descriptors can be found in Appendix D). FIM scoring consistently ranges from 0 -7, while SCIM III scoring options and descriptors vary between items. For all EFS items FIM scores were reduced and in some cases SCIM III scores were reduced in order to align item descriptions between instrument scoring. The highest possible number of EFS scores is equal to the lowest number of possible scores for either FIM or SCIM III for a given item, although in some cases further reduction was required to align score descriptors.

Exact FIM/SCIM III score equivalents or reductions are displayed in Figure 13 and are summarized below.

- Scores for feeding/eating, bathing, dressing, grooming, use of toilet, and stairs were re-scored to four possible scores broadly described as 0=Total Assistance, 1= Partial Assistance, 2=Independent with equipment or set-up, 3= Independent without equipment or set-up.
    - SCIM III feeding, bathing, grooming, and stairs consist of four scoring options thus SCIM scoring categories were not reduced as they were equivalent to the number of EFS categories (Table 9, Figure 13).
    - SCIM III dressing and use of toilet (originally five items), were reduced to four items to align score descriptors. A detailed example of item score reduction is show in Figure 14. SCIM III use of toilet contains five score options and differentiates partial assistance with and without cleaning. FIM does not specifically differentiate based on cleaning (it is one of several activities contributing to the percent of the activity an individual performs) for partial assistance thus these SCIM III categories were collapsed to better align with FIM (Table 9, Figure 13, Figure 14).
- SCIM III transfer scoring (bed, tub/toilet) consisted of three options, thus the EFS was reduced to three options; 0= Total assistance, 1= Partial assistance with or without equipment and, 2 = independent without equipment (Table 9, Figure 13).
- SCIM III mobility moderate distance was reduced from nine to five scoring options (Table 9, Figure 13).

In some cases, individual items were combined to form the EFS, such as with bathing where FIM has one item and SCIM III separates bathing into upper and lower extremity. Where items were combined to form a single EFS item, a score of 0 (total assistance) required both items, in this case

upper and lower extremity bathing to be scored 0 reflecting total assistance for both items. Similarly, for an original score of 4 reflecting independence without an assistive device, both upper and lower body bathing had to be scored as 4, reflecting independence in both items (Table 9). For an EFS score of 1 (partial assistance), either upper or lower body required partial assistance and for an EFS score of 2 (independent with equipment) one item had to be independent with equipment, while the second item could be the same or higher (independent without equipment).

In addition to bathing, this strategy was used for toilet and tub/shower transfers as FIM separates these items into toilet transfers, and tub or shower transfers, while SCIM III has only one item encompassing both of these transfers (Table 9).

The walking /wheeling item was the most challenging due to differences in FIM and SCIM III. Basic differences between FIM and SCIM III are that for the three SCIM III mobility items, each one encompasses wheelchair use and walking with varying distances and locations (indoor, moderate distance and outdoor), while FIM separates locomotion by wheelchair and walking with a functional modifier such that the most prevalent mode is used or both can be chosen, but only one score is assigned (Table 9). Figure 15 provides a detailed schematic demonstrating the crosswalk for walking/wheeling.

A number of items do not have an equivalent category in both measures, such as mobility in bed and action to prevent bed sores, wheelchair to car and ground to wheelchair transfers, which are only present in SCIM III.

In total, for the sixteen SCIM III mobility items, five were not used (mobility in bed and action to prevent bed sores, mobility- indoors and outdoors on even surface, transfers w/c to car, transfers ground to w/c) and 11 were retained. All FIM items were used

Table 9: SCIM and FIM expert panel conversions for all voluntary motor items and scores

SCIM	Expert Panel FIM SCIM (EFS)	FIM
Feeding:0-3	Eating	Eating:0-7
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance SCIM=1, FIM=2-4 2 = Independent with equipment or set-up SCIM=2, FIM=5,6 3 = Independent w/o equipment or set-up SCIM=3, FIM=7	
Bathing upper body:0-3 Bathing lower body:0-3	Bathing (upper and lower body)	Bathing:0-7
	0 = Total assistance SCIM (both upper body & lower body =0, FIM=0,1 1 = Partial assistance SCIM (upper & lower body) – 0 &1,2,3, 1&2, 1,2,3&0, 2&1 2 = Independent with equipment SCIM (upper & lower body) = 2&2, 2&3, 3&2, FIM=6 3 = Independent w/o equipment SCIM (both upper & lower body)=3, FIM=7	
Dressing upper body:0-4	Dressing upper body	Dressing upper body:0-7
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance SCIM=1, FIM =2-4 2 = Independent except fasteners or with equipment SCIM =2,3 FIM=5,6 3 = Independent w/o equipment SCIM=4, FIM=7	

Table 9 cont'd

SCIM	**Expert Panel FIM SCIM (EFS)**	**FIM**
Dressing lower body:0-4	Dressing lower body	Dressing lower body: 0-7
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance SCIM=1, FIM=2-4 2 = Independent except fasteners or with equipment SCIM =2,3 FIM=5,6 3 = Independent w/o equipment SCIM=4, FIM=7	
Grooming:0-3	Grooming	Grooming:0-7
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance SCIM=1, FIM=2-5 2 = Independent with equipment SCIM=2, FIM=6 3 = Independent w/o equipment SCIM=3, FIM=7	
Use of toilet:0,1,2,4,5	Using toilet	Toileting:0-7
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance SCIM=1,2 FIM=2-5 2 = Independent with equipment SCIM=4, FIM=6 3 = Independent w/o equipment SCIM=5, FIM=7	
Mobility in bed and action to prevent bed sores:0,2,4,6		No equivalent FIM item

Table 9 con't

SCIM	Expert Panel FIM SCIM (EFS)	FIM
Transfers bed-wheelchair 0-2	Transfers: bed-w/c	Transfers bed, chair, w/c:0-2
	0 = Total assistance SCIM=0, FIM=0,1 1 = Partial assistance or with equipment SCIM=1, FIM=2-6 2 = Independent w/o equipment SCIM=2, FIM=7	
Transfers w/c, toilet, tub:0-2	Transfers: toilet, tub/shower	Transfers toilet:0-7 Transfers tub or shower:0-7
	0 = Total assistance SCIM=0, FIM (both toilet and tub/shower transfers) =0,1 1 = Partial assistance or with equipment SCIM=1, FIM (toilet and shower tub/transfers) = 0&2-7,1&2-7, 2&1-7, 3&1-7, 4&1-7, 5&1-7, 6&1-7, 7&1-6, 0-7& 2, 0-7&3, 0-7&3, 0-7&4, 0-7&5, 0-7&6 2 = Independent w/o equipment SCIM=2, FIM (both toilet and tub/shower transfers) =7	
Mobility (indoors and outdoors on even surface) indoors:0-8	Not used	
Mobility (indoors and outdoors on even surface) for moderate distances(10-100 meters):0-8	Wheelchair/walk	Locomotion wheelchair:0-7 Locomotion walk:0-7
	0 = Total assistance SCIM=0, FIM (w/c or walk) =0,1 1 = Uses w/c (with or w/o assistance) SCIM=1,2, FIM (w/c) =2-5 2 = Walks with assistance or supervision (with or w/o equipment) SCIM =3 FIM (walk): 2-5 3 =Walks Independently with equipment SCIM=4-7, FIM(walk)=6 4 =Walks Independently w/o equipment SCIM=8, FIM(walk)=7	
Mobility (indoors and outdoors on even surface) outdoors (more than 100 meters)	Not used	

Table 9 cont'd

SCIM	Expert Panel FIM SCIM (EFS)	FIM
Stair Management:0-3	Climbing stairs	Locomotion stairs:0-7
	0 = Does not do or Total assistance SCIM=0, FIM =0,1 1 = Partial assistance SCIM=1, FIM =2-5 2 = Independent with equipment SCIM=2, FIM= 6 3 = Independent w/o equipment SCIM=3, FIM= 7	
Transfers w/c to car:0-2		No equivalent item
Transfers ground to w/c:0,1		No equivalent item

w/c= wheelchair, w/o = without

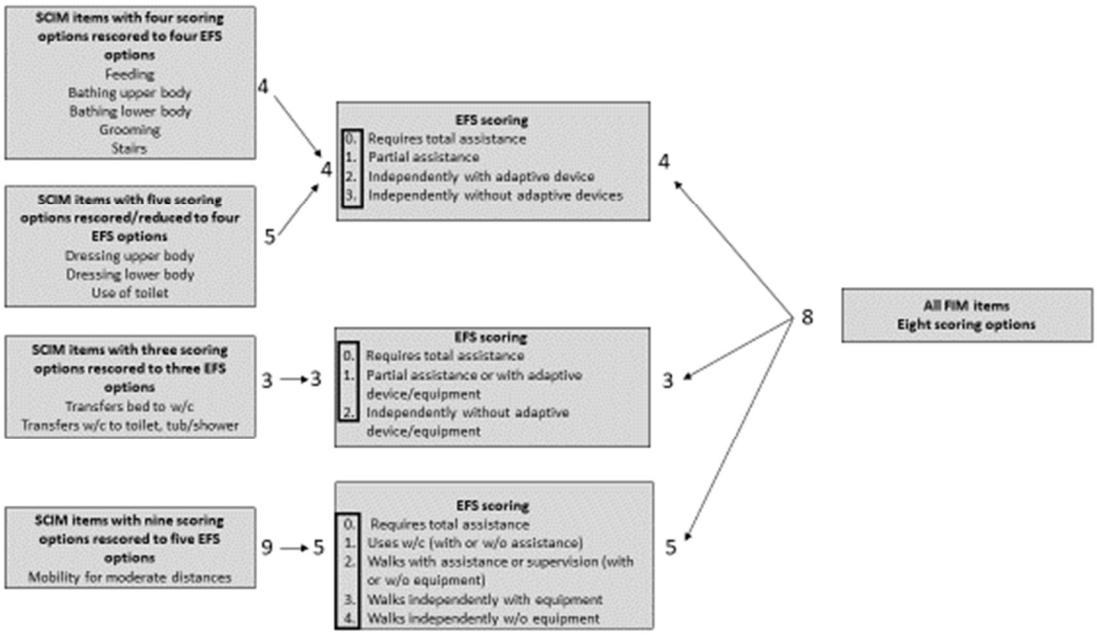


Figure 13: FIM/SCIM III score equivalents

SCIM III items and original scoring options are in the left column, with original FIM scoring options on the right. The equivalent Expert Panel FIM/SCIM III (EFS scoring) is in the middle column.

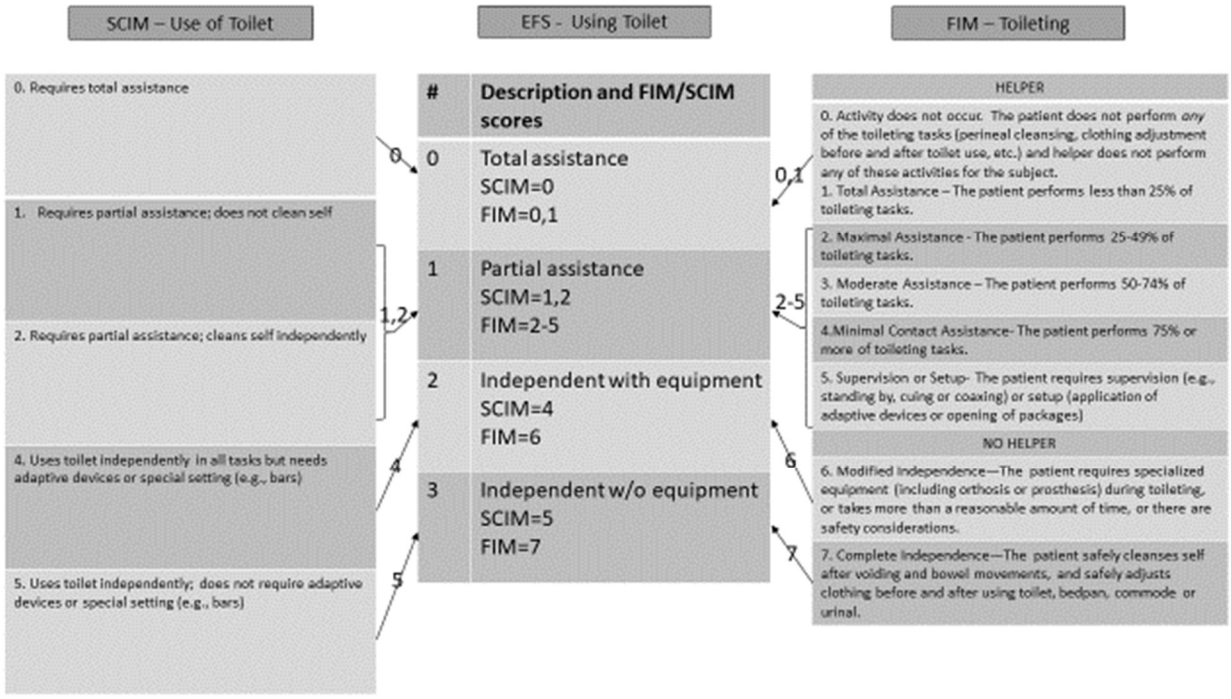


Figure 14: Example of item score collapsing for Expert Panel FIM/SCIM III using toilet item

Original SCIM III scores and descriptors for use of toilet are in the left hand column, with original FIM scores and descriptors on the right. Arrows indicate which scores were "matched" to the Expert Panel FIM/SCIM III (EFS) with descriptors in the middle column.

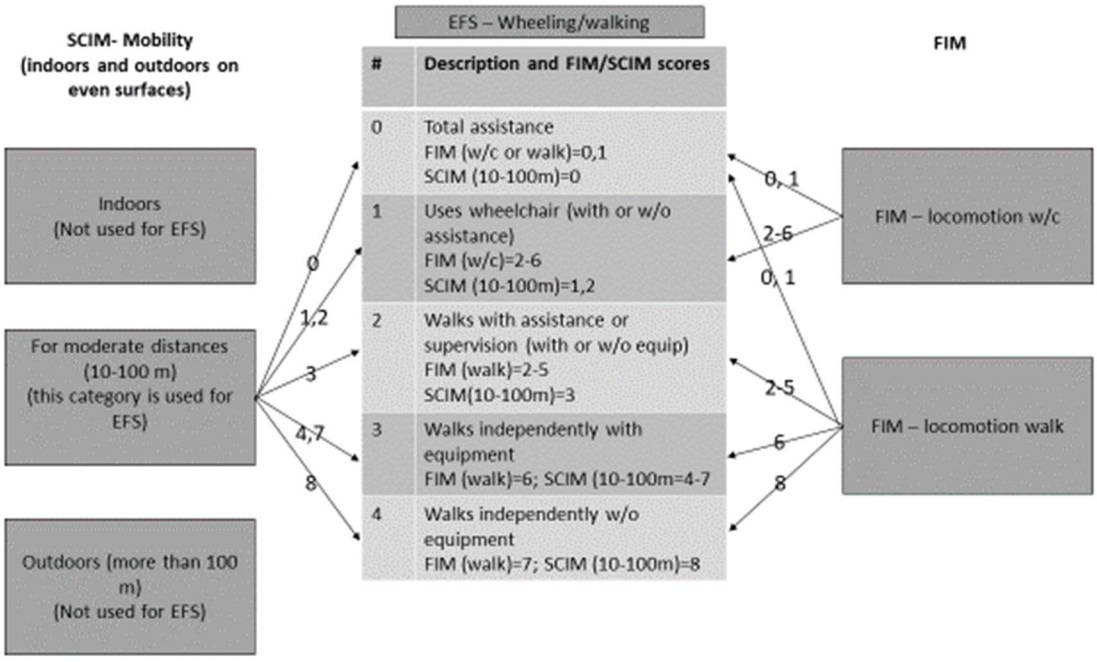


Figure 15: Expert panel wheeling/walking mobility item

SCIM III mobility items are displayed in the left column, while FIM locomotion items are displayed on the right. Arrows from items which were used in the Expert Panel FIM/SCIM III (EFS) and equivalent scores are indicated by arrows and in the middle column.

4.3.2 Method 2: Equipercentile crosswalk

Table 10 presents the crosswalk table for the FIM to EQ SCIM and Table 11 for the SCIM III to EQ FIM crosswalk. Total voluntary motor scores for FIM and SCIM III were rank ordered separately. Scores were then lined up side by side based on percentile rank. Figure 16 provides a graphical example of how the crosswalk tables are created for raw scores. In this graph a raw SCIM score of 28 is equivalent to a raw FIM score of 48, based on a percentile rank of approximately 60%. Score distributions are presented in Figure 17 and Figure 18. As both crosswalk score distributions were not irregular; no smoothing was required.

Table 10: Crosswalk table for FIM raw score to equipercentile SCIM conversion

FIM raw scores	Equipercentile SCIM conversion		FIM raw scores	Equipercentile SCIM conversion
11	0.00		48	27.91
12	0.83		49	29.17
13	2.00		50	30.00
14	2.86		51	30.57
15	3.00		52	31.20
16	3.67		53	32.29
17	4.00		54	33.40
18	5.00		55	34.25
19	5.50		56	35.00
20	6.17		57	35.00
21	7.08		58	36.00
22	8.00		59	36.67
23	8.75		60	37.00
24	9.74		61	37.38
25	10.93		62	38.00
26	11.45		63	38.67
27	12.08		64	39.71
28	13.00		65	40.45
29	13.40		66	41.92
30	14.25		67	44.63
31	15.00		68	47.07
32	15.94		69	50.27
33	17.11		70	53.93
34	18.00		71	57.13
35	18.00		72	58.43
36	18.90		73	60.40
37	19.46		74	61.20
38	20.00		75	63.80
39	20.91		76	64.29
40	21.00		77	65.00
41	22.00			
42	22.67			
43	23.60			
44	24.78			
45	25.75			
46	26.75			
47	27.00			

Table 11: Crosswalk table for SCIM raw score to equipercentile FIM conversion

SCIM raw scores	Equipercentile FIM conversion		SCIM raw scores	Equipercentile FIM conversion
0	11.10		37	59.94
1	12.00		38	62.07
2	13.00		39	63.20
3	15.06		40	64.55
4	16.33		41	65.38
5	18.36		42	66.00
6	19.67		43	66.00
7	20.86		44	67.00
8	22.14		45	67.00
9	23.33		46	67.80
10	24.06		47	68.00
11	25.30		48	68.00
12	26.69		49	68.67
13	28.40		50	69.00
14	29.67		51	69.00
15	30.92		52	69.50
16	32.00		53	70.00
17	32.80		54	70.00
18	34.47		55	70.00
19	36.44		56	70.50
20	37.72		57	71.00
21	39.38		58	71.67
22	41.30		59	72.00
23	42.40		60	73.00
24	43.25		61	73.67
25	44.22		62	74.00
26	45.14		63	75.00
27	46.89		64	75.56
28	48.00		65	76.67
29	48.83			
30	50.20			
31	51.67			
32	52.71			
33	53.82			
34	54.33			
35	56.53			
36	58.33			

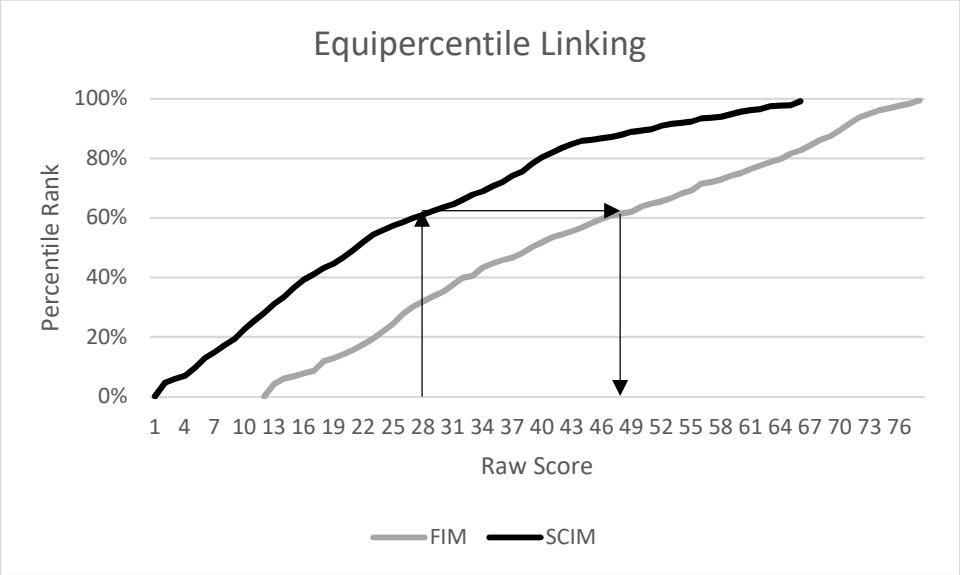


Figure 16: Equipercentile linking raw FIM to raw SCIM by percentile rank

For example, a raw SCIM III score of 28 is equivalent to a raw FIM score of 48, based on a percentile rank of approximately 60%.

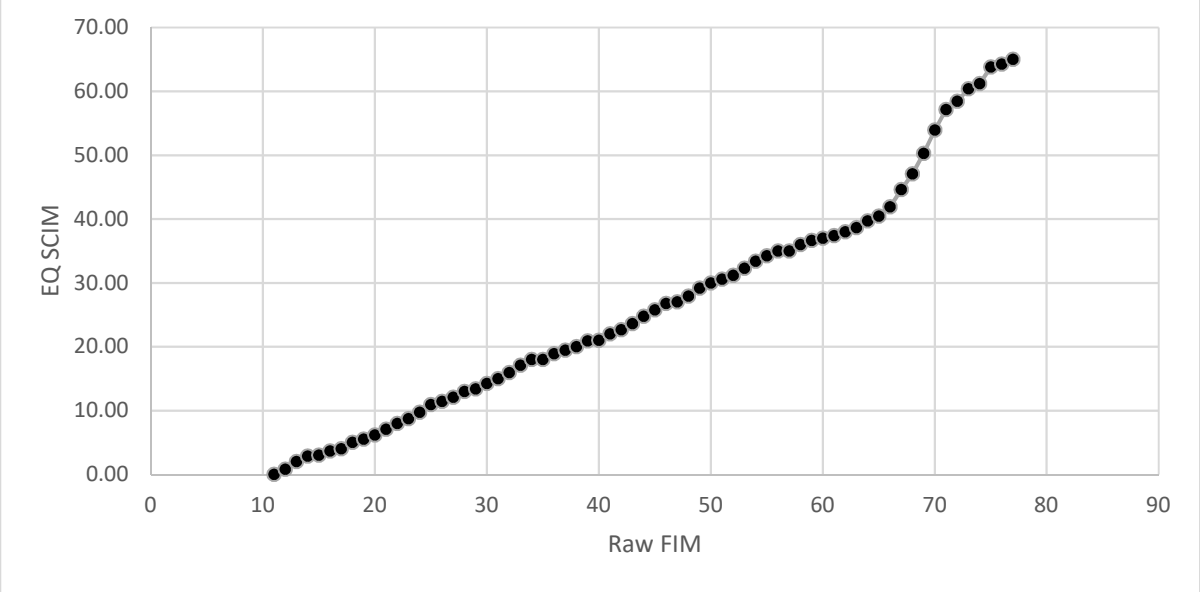


Figure 17: Equipercentile SCIM score distribution

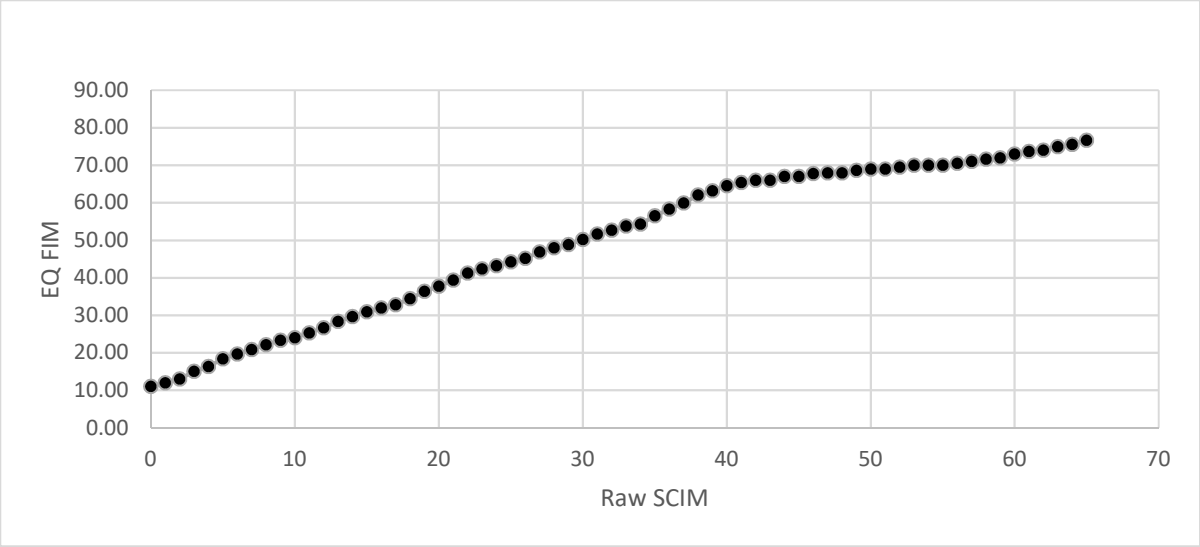


Figure 18: Equipercentile FIM score distribution

4.3.3 Method 3: Rasch crosswalk

Table 12 and Table 13 present the crosswalk tables for SCIM to Rasch FIM (R FIM) and FIM to Rasch SCIM (R SCIM). As with equipercentile, two crosswalks are presented due to differences in rounding. The person item map from which the crosswalk was created after co-calibration of the measures is demonstrated in Figure 19. The person item map displays the person's ability against item difficulty, by logits ("Measures" on the map). Both persons and items approximate a normal distribution with negligible floor and ceiling effects (3.4% and 0.2% respectively). There is some evidence of potential gaps in content, as there are no items at the lowest person ability levels. Exact logit difficulty is displayed in Table 14. In order to compare item difficulty between the scales, item functions are matched up for FIM and SCIM III items in Table 15. Item difficulty ranges from -2.31 to 2.28 (for all SCIM and FIM items together) with FIM stairs as the most difficult item and SCIM feeding as the easiest. The range of difficulty for FIM items is -2.05 to 2.28 (a total range of 4.33), while the range for SCIM items is -2.31 to 2.04 (a total range of 4.09). In most cases similar item functions in FIM have a similar item difficulty in

SCIM III. For FIM/SCIM III items with similar descriptors and 1:1 item matching (eat, dress upper body, dress lower body, grooming, use of toilet, transfer bed to wheelchair, stairs), differences between items were generally quite small (0.04- 0.67), with an average difference of 0.21 logits. Items with greater than 1 logit difference are bathing (FIM bathing and SCIM bathing upper body) and mobility items. As SCIM III separates bathing items into upper and lower body and FIM does not, this may explain the difference. Mobility items are very different in FIM and SCIM III with FIM using a functional modifier to only separate wheelchair vs. walking, while SCIM incorporates these differences within scores for each mobility item (lower scores=wheelchair, higher scores = walking).

Table 16 summarizes the psychometric properties of the combined FIM/SCIM II item bank, including person and item reliability and separation. Reliability in the context of Rasch analysis indicates if differences between individual participant abilities (or items) are due to actual differences in person ability (or items) or measurement error. Person separation indicates the number of strata or distinct functional groups, while item separation indicates strata within items. Reliability was high for both persons (0.94) and items (1.00). Six person strata or distinct functional categories were identified.

Table 12: Crosswalk table for FIM raw score to Rasch SCIM conversion

FIM Raw SCORE		Rasch SCIM conversion	FIM Raw Score		Rasch SCIM conversion
11		1	48		26
12		2	49		27
13		3	50		28
14		4	51		28.5
15		5	52		29
16		6	53		30
17		6	54		31
18		7	55		31.5
19		8	56		32
20		8	57		33
21		9	58		34
22		9	59		35
23		10	60		36
24		11	61		37
25		11	62		38
26		12	63		39
27		12.5	64		41
28		13	65		43
29		14	66		44
30		15	67		46
31		15	68		49
32		16	69		51
33		17	70		53
34		17	71		55
35		18	72		57
36		19	73		60
37		19	74		62
38		20	75		63
39		21	76		64
40		21	77		65
41		22			
42		23			
43		23			
44		24			
45		24.5			
46		25			
47		26			

Table 13: Crosswalk table for SCIM raw score to Rasch FIM conversion

SCIM Raw SCORE		Rasch FIM Conversion	SCIM Raw SCORE		Rasch FIM Conversion
0		11	37		61
1		11	38		62
2		12	39		63
3		13	40		63
4		14	41		64
5		15	42		65
6		17	43		65
7		18	44		66
8		20	45		66
9		22	46		67
10		23	47		67
11		25	48		68
12		26	49		68
13		28	50		69
14		29	51		69
15		31	52		70
16		32	53		70
17		33	54		70
18		35	55		71
19		36	56		71
20		38	57		72
21		39.5	58		72
22		41	59		73
23		42.5	60		73
24		44	61		74
25		46	62		74
26		47	63		75
27		49	64		76
28		50	65		77
29		52			
30		53			
31		54			
32		56			
33		57			
34		58			
35		59			
36		60			

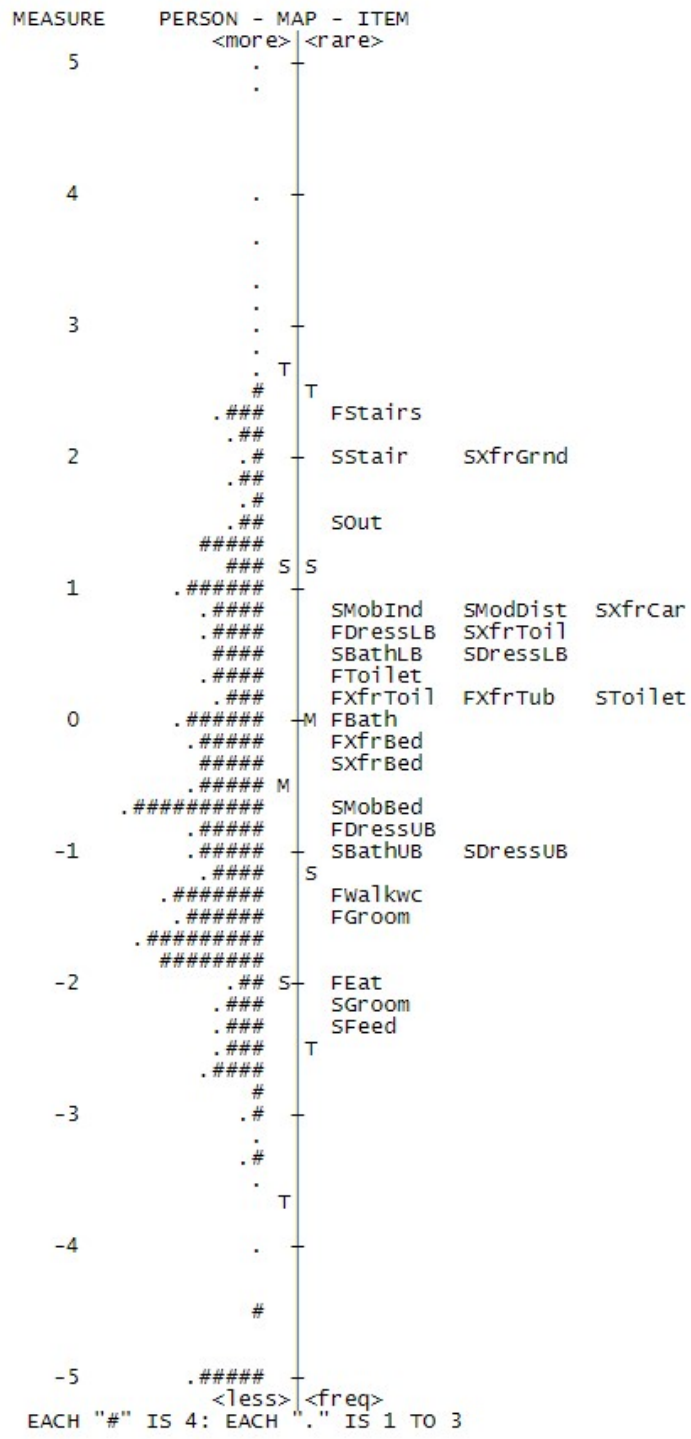


Figure 19: Rasch person-item map

This figure illustrates person ability and item difficulty on the same linear scale. Person ability is on the left side of the figure, with item difficulty on the right, with higher ability/ difficulty at the top and lower ability/difficulty at the bottom, as measured by logits (on the far left).

Table 14: Rasch item difficulty by logits for FIM and SCIM III items

Item difficulty in logits	Item
2.28	FIM stairs
2.04	SCIM stairs
1.96	SCIM transfer ground
1.46	SCIM mobility outdoors
0.92	SCIM transfer car
0.90	SCIM moderate distance
0.78	SCIM mobility indoors
0.66	SCIM transfer toilet
0.59	FIM dress lower body
0.48	SCIM bath lower body
0.44	SCIM dress lower body
0.30	FIM use of toilet
0.24	FIM transfer tub or shower
0.22	SCIM use of toilet
0.12	FIM transfer toilet
0.03	FIM bathing
-0.24	FIM transfer bed
-0.28	SCIM transfer bed
-0.60	SCIM mobility bed
-0.87	FIM dress upper body
-0.92	SCIM dress upper body
-1.06	SCIM bath upper body
-1.26	FIM walk/wheelchair
-1.57	FIM Groom
-2.05	FIM Eat
-2.24	SCIM Groom
-2.31	SCIM Feed

Table 15: Comparison of item difficulty by logits for FIM and SCIM III item functions

Item function	FIM item(s) and logits	logits	SCIM item(s) and logits	logits	Difference in logits (negative value indicates SCIM is more difficult)
Eat/Feed	eat	-2.05	feed	-2.31	0.26
Bathing	bathing	0.03	bathing upper body	-1.06	1.09
			bathing lower body	0.48	-0.45
Dressing	dressing upper body	-0.87	dressing upper body	-0.92	0.05
	dressing lower body	0.59	dressing lower body	0.44	0.15
Grooming	grooming	-1.57	grooming	-2.24	0.67
Use of toilet	toileting	0.3	use of toilet	0.22	0.08
Mobility in bed	No FIM item		mobility in bed	-0.6	Not applicable
Transfer bed to wheelchair	transfer bed to wheelchair	-0.24	transfer bed, chair, wheelchair	-0.28	0.04
Transfer toilet, tub/shower	transfer toilet	0.12	transfer wheelchair, toilet, tub	0.66	-0.54
	transfer tub or shower	0.24			-0.42
Mobility	walk/wheelchair	-1.26	mobility indoors	0.78	-2.04
			mobility moderate distances	0.9	-2.16
			mobility outdoors		-2.72
Stairs	locomotion stairs	2.28	stair management	2.04	0.24
Transfers w/c to car	No FIM item		transfers wheelchair to car	0.92	Not applicable
Transfers ground to wheelchair	No FIM item		transfers ground to wheelchair	1.96	Not applicable

Table 16: Rasch psychometric properties for the combined FIM/SCIM III item bank

FIM-SCIM III item bank	
Person reliability	0.94
Person separation index	3.93
Person strata	5.57
Person ability (logits)	mean = -0.64, SD = 1.84, max= 6.05, min = -5.75 (range 11.8)
Item difficulty (logits)	mean = .00, SD = 1.23, max= 2.28, min=-2.31 (range 4.59)
Misfitting items (high fit, no items with low fit) (percent,n)	7% (2/27)
Floor effect (percent,n)	3.4% (23/662)
Ceiling effect (percent,n)	0.2% (1/662)

SD= standard deviation, max=maximum, min=minimum

4.4 Primary crosswalk assessment

The primary assessment of the strength of a crosswalk is the correlation coefficient. Correlations for all three methods in all three databases as well as the associated degree (percent) of uncertainty, are presented in Table 17. Primary correlations (presented in bold) for EFS are between EFS FIM and EFS SCIM, while correlations for equipercentile and Rasch are between raw and crosswalked scores. As noted, correlations should be greater than 0.866 to reduce uncertainty by at least 50%.[8]

Correlations for raw scores were high across all databases (0.898 – 0.961). These high correlations suggest that FIM and SCIM III are measuring a similar construct. Additional correlations are presented as well (not bold). SCIM and FIM raw scores to respective FIM and SCIM EFS scores are high across all datasets (0.968-0.984) indicating that despite the collapsing of items and scores, the relationship between raw and crosswalked data remains high. EQ FIM to EQ SCIM correlations and R FIM to R SCIM correlations are also high (0.891 – 0.955) indicating the relationship between the final crosswalked FIM and SCIM III measures remains strong.

Assessment by method

Primary comparisons were high across the development and validation databases for all methods, exceeding 0.866 (0.897-0.971). The degree of uncertainty ranged from 56 % to 77%. No method had consistently higher correlations than another.

Assessment by dataset

Correlations were high across all datasets. On average RHSCIR correlations were highest (0.970) and lowest across SWISS (0.910), with Anderson in the middle (0.9244).

Assessment by outcome measure (FIM/SCIM)

Raw SCIM to crosswalked SCIM III correlations were slightly lower than raw FIM to crosswalked FIM correlations.

Table 17: Pearson correlation coefficients for all methods with degree of uncertainty for primary correlations

	SWISS (n=662)	**Anderson (n=119)**	**RHSCIR (n=133)**
SCIM and FIM total voluntary motor scores – raw scores (r)	0.898	0.922	0.961
Equilibrated FIM/SCIM - Method 1			
EFS FIM and EFS SCIM [r, degree of uncertainty as a percentage]	**0.911 (59%)**	**0.920 (61%)**	**0.972 (77%)**
SCIM raw scores EFS SCIM (r)	0.980	0.968	0.984
FIM raw scores and EFS FIM (r)	0.979	0.971	0.982
Equipercentile FIM/SCIM – Method 2			
SCIM raw scores and EQ SCIM (r, degree of uncertainty as a percentage)	**0.901 (57%)**	**0.921 (61%)**	**0.969 (75%)**
FIM raw scores and EQ FIM (r, degree of uncertainty as a percentage)	**0.918 (60%)**	**0.931 (64%)**	**0.971 (76%)**
EQ FIM and EQ SCIM (r)	0.903	0.929	0.955
Rasch FIM/SCIM – Method 3			
SCIM raw scores and R SCIM (r, degree of uncertainty as a percentage)	**0.897 (56%)**	**0.917(60%)**	**0.966 (74%)**
FIM raw scores and R FIM (r, degree of uncertainty as a percentage)	**0.917 (60%)**	**0.933 (64%)**	**0.971 (76%)**
R FIM to R SCIM (r)	0.891	0.917	0.944

Correlations reported to three decimal places in order to compare with recommended value exceeding 0.866.
r= Pearson correlation, EFS= expert panel FIM/SCIM III, EQ= equipercentile, R=Rasch
Bold correlations are primary correlations.

Specific Aim 2: A crosswalk/crosswalks for FIM and SCIM III voluntary motor function items will be created using three conceptually different methods: expert panel linking, equipercentile linking and Rasch analysis co-calibration. Correlations between actual and cross-walked scores using the crosswalk(s) for each of the three methods will be assessed.

Hypothesis: Correlations will exceed established criteria (0.866)[8] using the crosswalk(s) for at least one of the three methods.

Null Hypothesis: Correlations will not exceed established criteria (0.866)[8] using the crosswalk(s)

Specific Aim 3: Validate the three crosswalk methods in a separate dataset.

Hypothesis: Correlations in the validation dataset will exceed established criteria (0.866) using crosswalk(s) for each of the three methods.

Null Hypothesis: Correlations in the validation dataset will not exceed established criterion (0.866) using crosswalk(s), for each of the three methods.

As correlation coefficients for all three methods were greater than 0.866 in both the development and validations databases, the hypotheses for Aims 1 and 2 were accepted and the null hypotheses rejected. As all correlations exceeded 0.866, additional criteria (score distribution, distribution of differences- Bland Altman, amount of difference between distributions – Cohen's effect size, point differences and subgroup invariance) were assessed for all methods.

4.5 Secondary crosswalk assessments

All assessments compare EFS SCIM and EFS FIM and raw to crosswalked scores for SCIM III and FIM.

4.5.1 Score distributions

Score distributions for all data sets are displayed in Figure 21, Figure 22, and Figure 23. The primary assessment of score distribution was by visual comparison of graphs. In addition, in Table 18, differences in means, means within one standard deviation of one another and standard deviations within one unit of each other, and overlap of confidence intervals for skewness and kurtosis were examined numerically as per Ketchum et al.[125]

Assessment by method

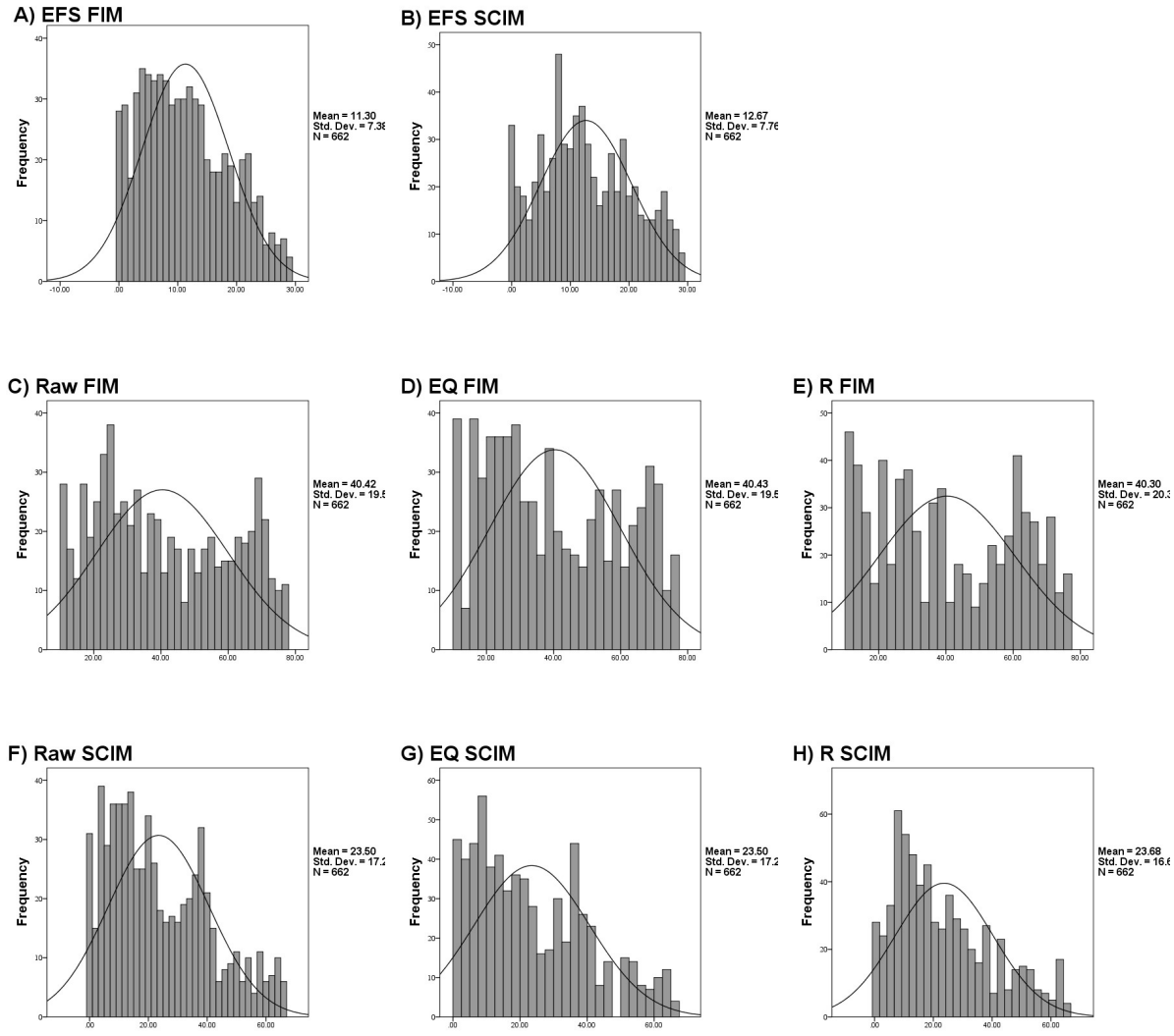
Score distributions for all three methods were similar with no striking differences between EFS FIM and EFS SCIM, raw and crosswalked equipercentile and Rasch scores. In examining the objective criteria suggested by Ketchum et al.[125], the means for the expert panel method were significantly different within all datasets, and EFS FIM means consistently were lower than EFS SCIM means. All other significant differences between means for raw and crossswalked scores occurred in the Anderson dataset. Significant differences were found between raw score to EQ SCIM, RFIM and RSCIM. All of these comparisons also demonstrated standard deviations with greater than one-unit difference, except SCIM raw scores to EQ SCIM.

Assessment by dataset

Other than EFS comparisons, all comparisons that did not meet criteria occurred in the Anderson dataset (raw FIM to R FIM, raw SCIM to EQ SCIM and R SCIM).

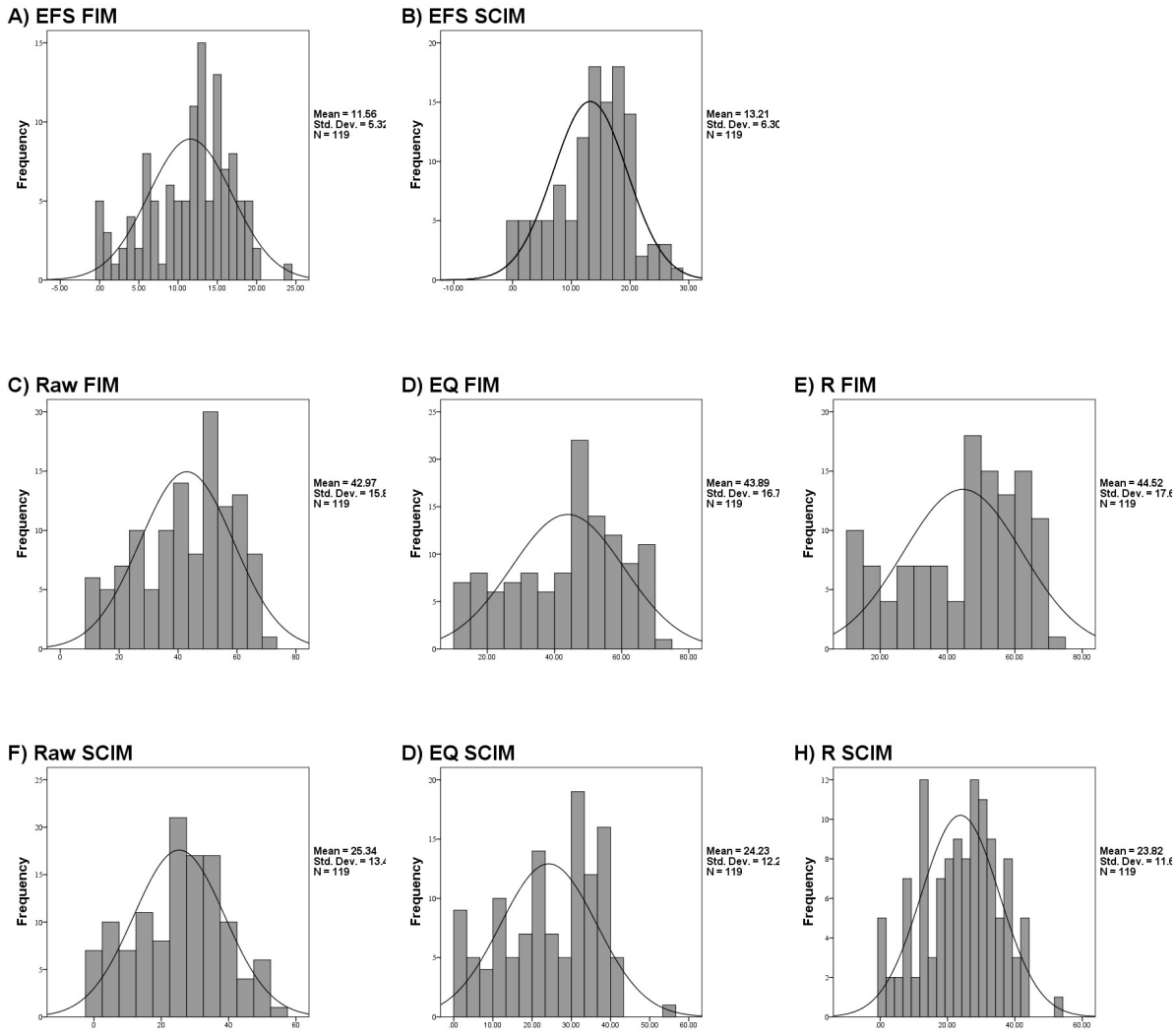
Assessment by outcome measure (FIM/SCIM)

In the SWISS and RHSCIR datasets, SCIM is skewed to the left, possibly indicating a floor effect. The expert panel method across all databases resulted in lower EFS FIM than EFS SCIM scores. As noted above, FIM scores are consistently lower than SCIM scores for the expert panel method. Significant differences were found between means in at least one dataset for all comparisons except FIM raw scores to EQ FIM.



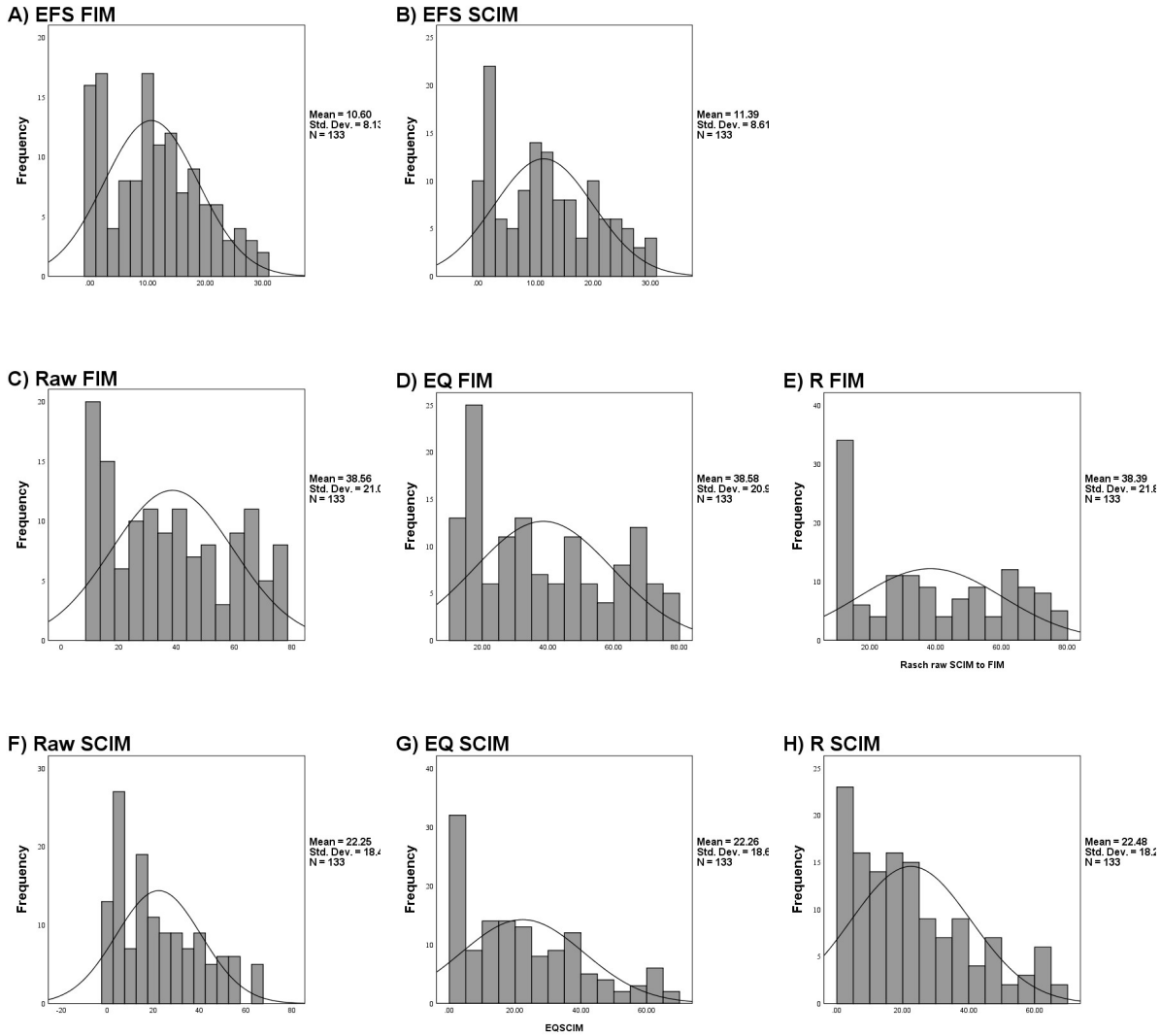
A) expert panel FIM/SCIM III (EFS) FIM, B) expert panel FIM/SCIM III (EFS) SCIM, C) raw FIM, D) equipercentile (EQ) FIM, E) Rasch (R) FIM, F) raw SCIM, G) equipercentile (EQ) SCIM, H) Rasch (R) SCIM

Figure 20: **Score distributions SWISS dataset**



A) expert panel FIM/SCIM III (EFS) FIM, B) expert panel FMm/SCIM III (EFS) SCIM, C) raw FIM, D) equipercentile (EQ) FIM, E) Rasch (R) FIM, F) raw SCIM, G) equipercentile (EQ) SCIM, H) Rasch (R) SCIM

Figure 21: Score distributions Anderson dataset



A) expert panel FIM/SCIM III (EFS) FIM, B) expert panel FIM/SCIM III (EFS) SCIM, C) raw FIM, D) equipercentile (EQ) FIM, E) Rasch (R) FIM, F) raw SCIM, G) equipercentile (EQ) SCIM, H) Rasch (R) SCIM

Figure 22: Score distributions RHSCIR dataset

Table 18: Score distribution (mean, standard deviation, skewness and kurtosis) for each method and dataset

(Comparisons are within datasets -bold columns)

	SWISS		**Anderson**		**RHSCIR**	
	EFS FIM	EFS SCIM	EFS FIM	EFS SCIM	EFS FIM	EFS SCIM
Mean	**10.60***	**11.39***	**11.56***	**13.21***	**10.60***	**11.39***
SD	8.13	8.62	5.32	6.30	8.13	8.62
Skewness	0.40	0.35	-0.49	-0.42	0.40	0.35
Kurtosis	-0.75	-0.96	-0.39	0-.30	-0.75	-0.96
	FIM raw scores	EQ FIM (converted)	FIM raw scores	EQ FIM	FIM raw scores	EQ FIM
Mean	40.42	40.43	42.97	43.89	38.56	38.58
SD	19.54	19.53	15.88	16.74	21.09	20.97
Skewness	0.23	0.32	-0.42	-0.40	0.26	0.29
Kurtosis	-1.24	-1.24	-0.87	-0.83	-1.23	-1.29
	FIM raw scores	RFIM	FIM raw scores	RFIM	FIM raw scores	RFIM
Mean	40.42	40.30	**42.97***	**44.52***	38.56	38.39
SD	19.54	20.35	**15.88****	**17.63****	21.09	21.82
Skewness	0.23	0.15	-0.42	-0.55	0.26	0.22
Kurtosis	-1.24	-1.34	-0.87	-0.89	-1.23	-1.39
	SCIM raw scores	EQ SCIM	SCIM raw scores	EQ SCIM	SCIM raw scores	EQ SCIM
Mean	23.50	23.50	**25.34***	**24.23***	22.25	22.26
SD	17.21	17.20	13.49	12.25	18.44	18.63
Skewness	0.62	0.62	-0.18	-0.34	0.64	0.68
Kurtosis	-0.53	-0.53	-0.67	-0.78	-0.63	-0.46
	SCIM raw scores	RSCIM	SCIM raw scores	RSCIM	SCIM raw scores	RSCIM
Mean	23.50	23.68	**25.34***	**23.82***	22.25	22.48
SD	17.21	16.68	**13.49****	**11.63****	18.44	18.21
Skewness	0.62	0.72	-0.18	-0.15	0.64	0.75
Kurtosis	-0.53	-0.41	-0.67	-0.59	-0.63	-0.35

SD= standard deviation, EFS= expert panel FIM/SCIM III, EQ=equipercentile, R=Rasch

*Significant difference between means (<0.05)

**Standard deviation > 1 unit difference

4.5.2 Bland Altman

Bland Altman plots were visually assessed for distribution of differences (Figure 23, Figure 24 and Figure 25) for all comparisons by graphing differences in scores against means for comparisons of

raw and crosswalked scores. EFS FIM and EFS SCIM were also plotted for comparison. Outliers were analyzed as summarized in "Additional Analyses" below.

Assessment by method

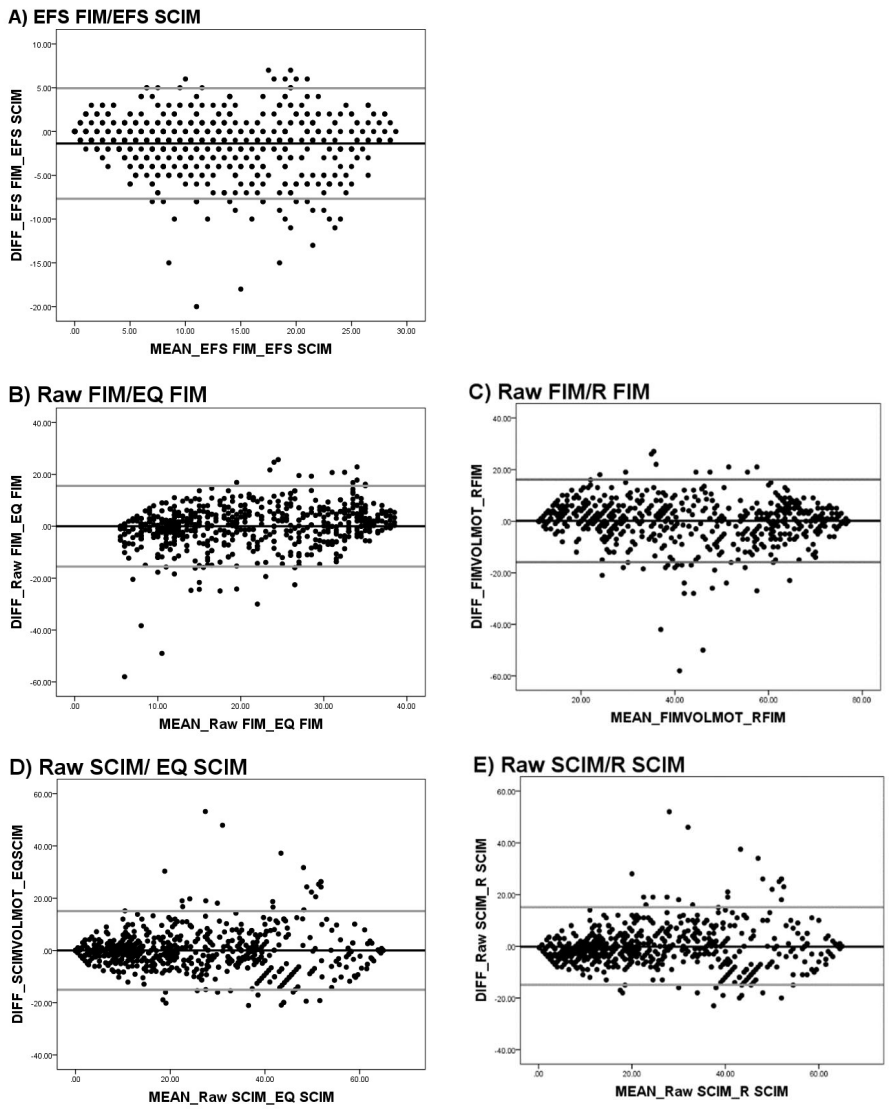
No clear patterns are apparent when comparing methods.

Assessment by dataset.

No clear patterns are apparent when comparing datasets.

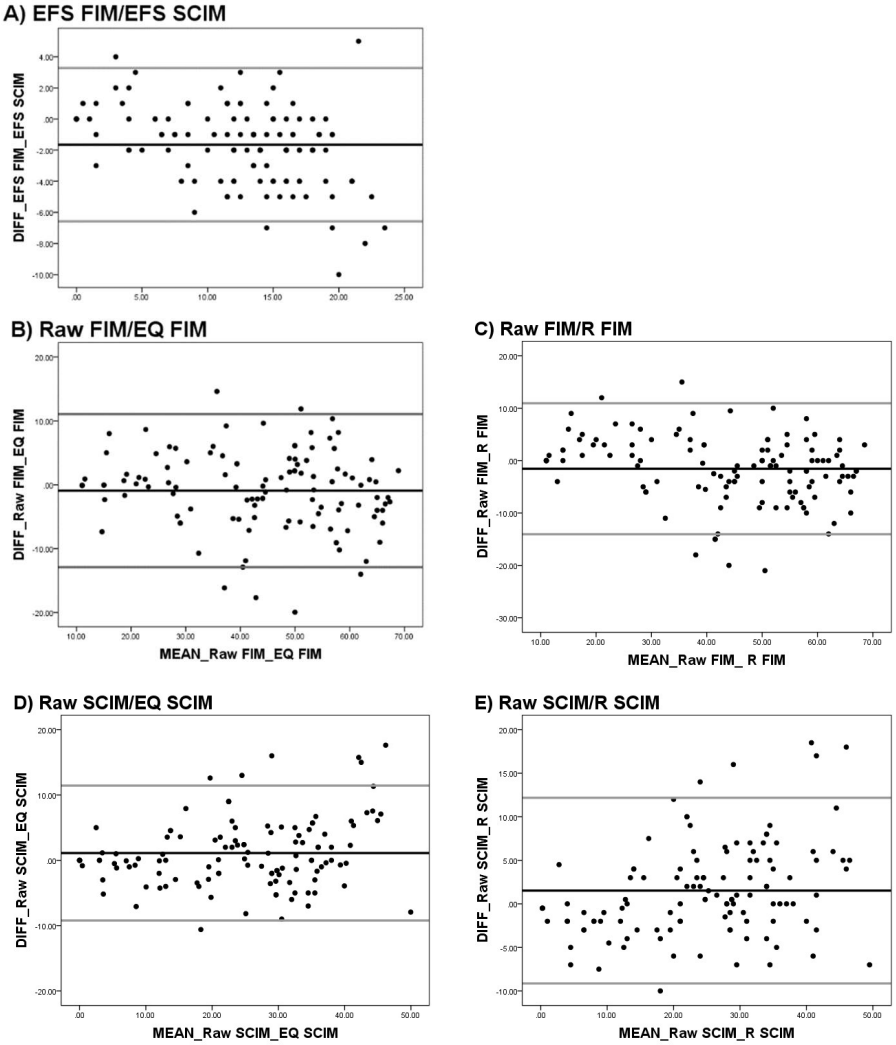
Assessment by outcome measure (FIM/SCIM)

Across the Anderson and RHSCIR datasets, FIM and particularly SCIM III have less variability at the low end of the scale, with larger differences at the high end of the scale.



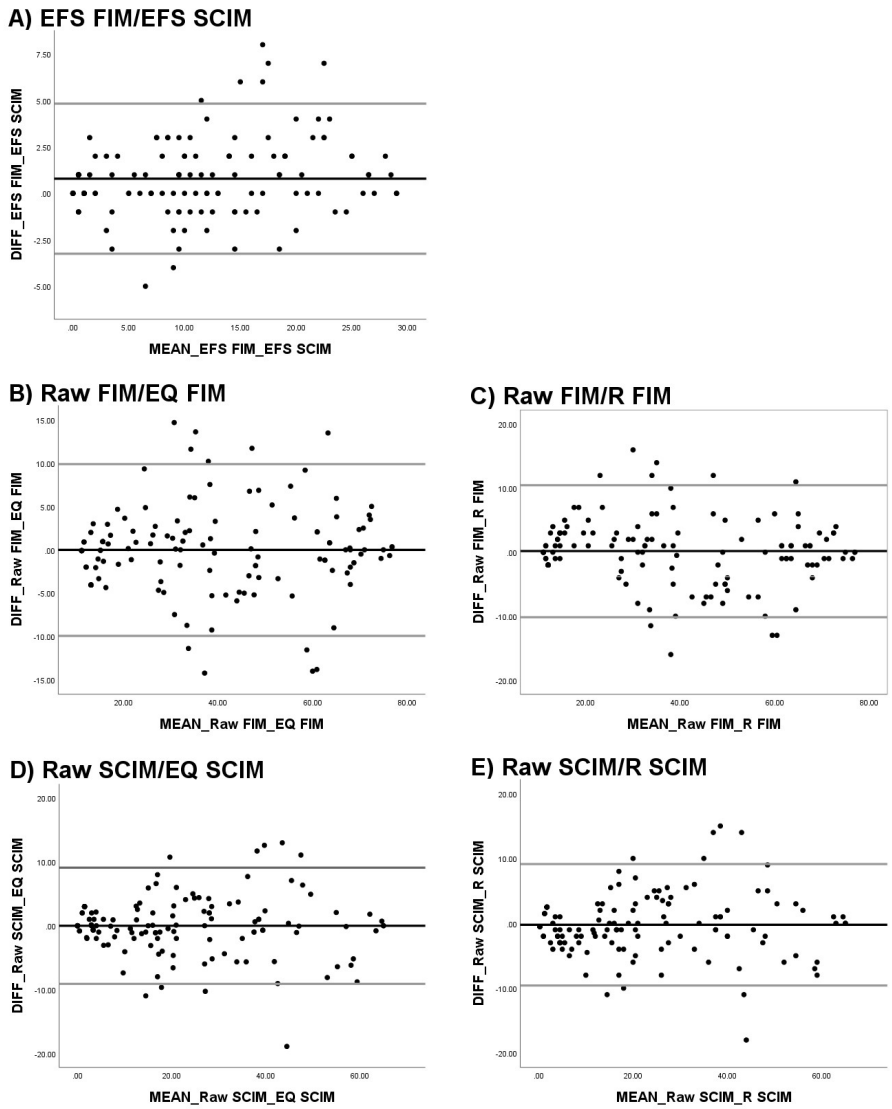
Plot of mean score against difference in scores for A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 23: Bland Altman plots SWISS dataset



Plot of mean score (x axis) against difference in scores for A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 24: Bland Altman plots Anderson dataset



Plot of mean score (x axis) against difference in scores for A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 25: Bland Altman plots RHSCIR dataset

4.5.3 Cohen's effect size

Cohen's effect sizes are displayed in Table 19. All effect sizes were expected to be small (0.20 or less). The smaller the effect size, the greater the overlap between data. For example, with an effect size

of 0, the overlap in distributions is 100%, 92% at .20, and 80.3% at 0.50. Negligible (< 0.20) effect sizes were noted for most comparisons.

Assessment by method

Medium and large effect sizes were noted for EFS FIM to EFS SCIM comparisons across all data sets (SWISS= 0.42, Anderson -0.65, RHSCIR = 0.38).

Assessment by dataset

Effect sizes larger than 0.20 were noted in the Anderson dataset for all comparisons except raw FIM and EQ FIM comparison. Most effect sizes only slightly exceeded 0.20 (raw FIM to R FIM [- 0.24] raw SCIM to EQ SCIM [.21] and R SCIM [0.28], except the EFS FIM and EFS SCIM which was 0.65. This is between a medium and large effect size, which results in a 74.5% overlap in distributions.

Assessment by outcome measure (FIM/SCIM)

Effect sizes slightly exceeded 0.20 in at least one dataset for all comparisons except FIM raw scores and EQ FIM comparisons.

Table 19: Cohen's effect size for each method and dataset

Method	SWISS	Anderson	RHSCIR
EFS FIM and EFS SCIM	**0.42***	**0.65 ***	**0.38 ***
FIM raw scores and EQ FIM	0.00	0.15	0.00
FIM raw scores and R FIM	0.02	**0.24 ***	0.03
SCIM raw scores and EQ SCIM	0.00	**0.21 ***	0.00
SCIM raw scores and RSCIM	0.02	**0.28***	0.05

EFS= expert panel FIM/SCIM III, EQ = equipercentile, R=Rasch
*Items exceed a small effect size of 0.20

4.5.4 Point Differences

Point differences between raw and crosswalk scores are illustrated in Table 20. The distribution of point differences in shown in Figure 26, Figure 27and Figure 28 for SWISS, Anderson and RHSCIR respectively. In looking at distributions, no striking differences are noted for assessment by method, dataset or outcome measure.

Assessment by method

The pre-set cut off was achieved for all methods except EFS FIM and EFS SCIM comparisons, which was 74%. The expert panel had the lowest percent of point differences within 0.5 standard deviations across all datasets at 81%. All other methods were between 88-89%.

Assessment by dataset

The pre-set cut-off was achieved in all datasets except for one comparison in Anderson. RHSCIR had the highest percentage of scores within 0.5 standard deviations (93%), followed by Anderson (87%) and SWISS (82%).

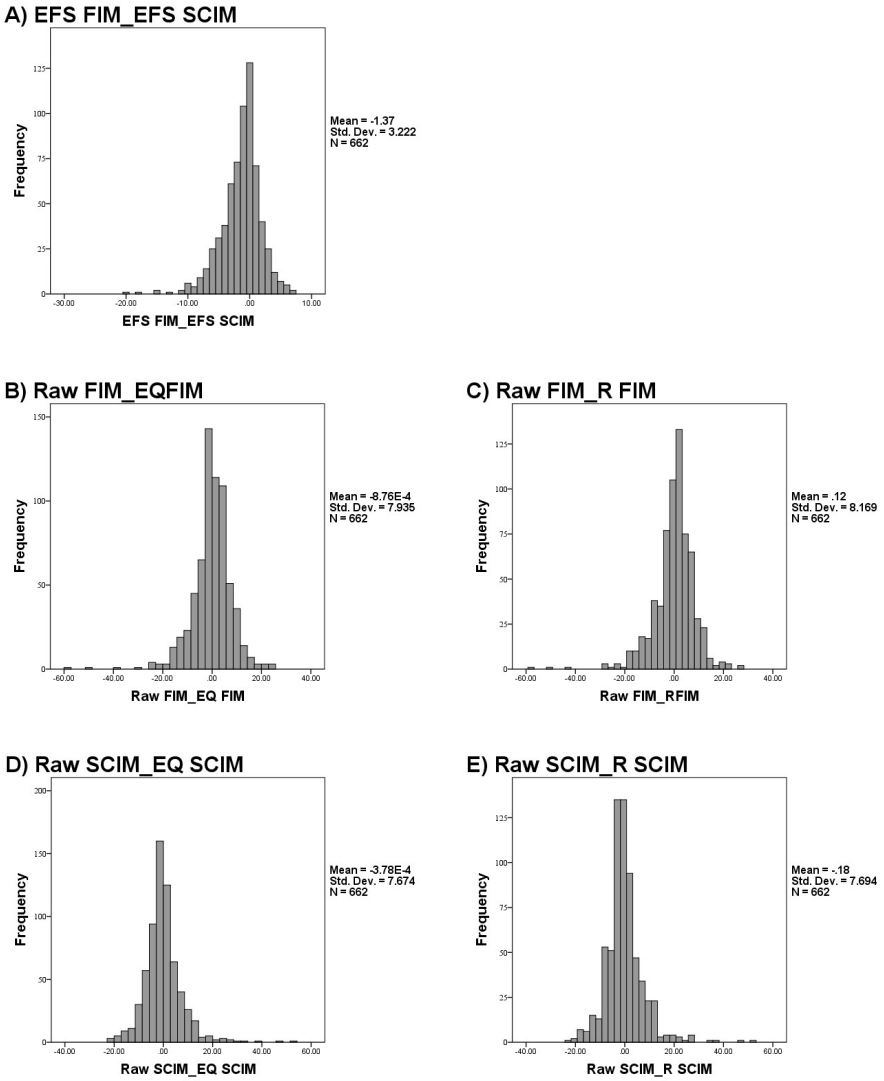
Assessment by outcome measure (FIM/SCIM)

No differences were noted between outcome measures.

Table 20: Percent point differences less than 0.5 standard deviation from the mean for each method and dataset

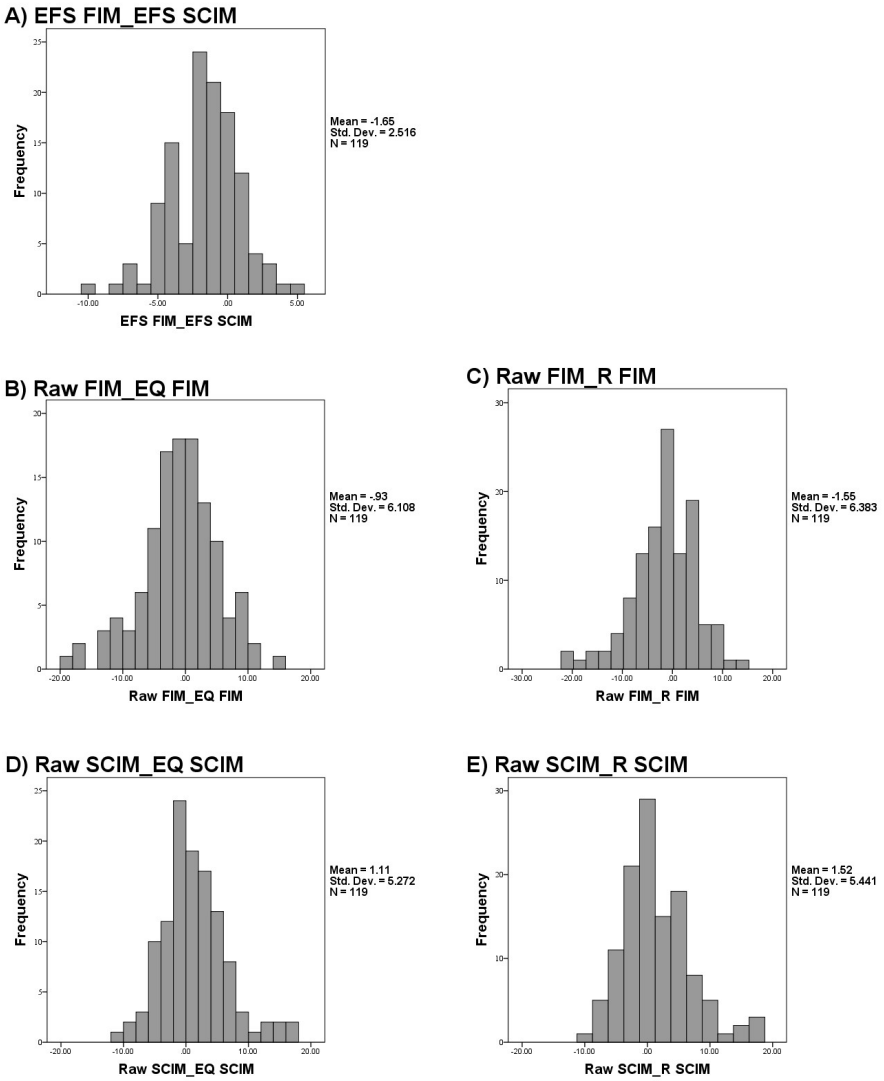
Method	SWISS	Anderson	RHSCIR
EFSFIM and EFSSCIM	77.7%	**74.0%**	91.7%
FIM VOL MOTOR SUM and EQFIM	83.5%	89.9%	93.2%
FIM VOL MOTOR SUM and RFIM	84.1%	89.1%	93.2%
SCIM VOL MOTOR SUM and EQ SCIM	81.7%	91.6%	94.0%
SCIM VOL MOTOR SUM and RSCIM	82.2%	90.7%	91.7%

EFS= expert panel FIM/SCIM III, EQ = equipercentile, R=Rasch
Items in bold are below the 75% threshold



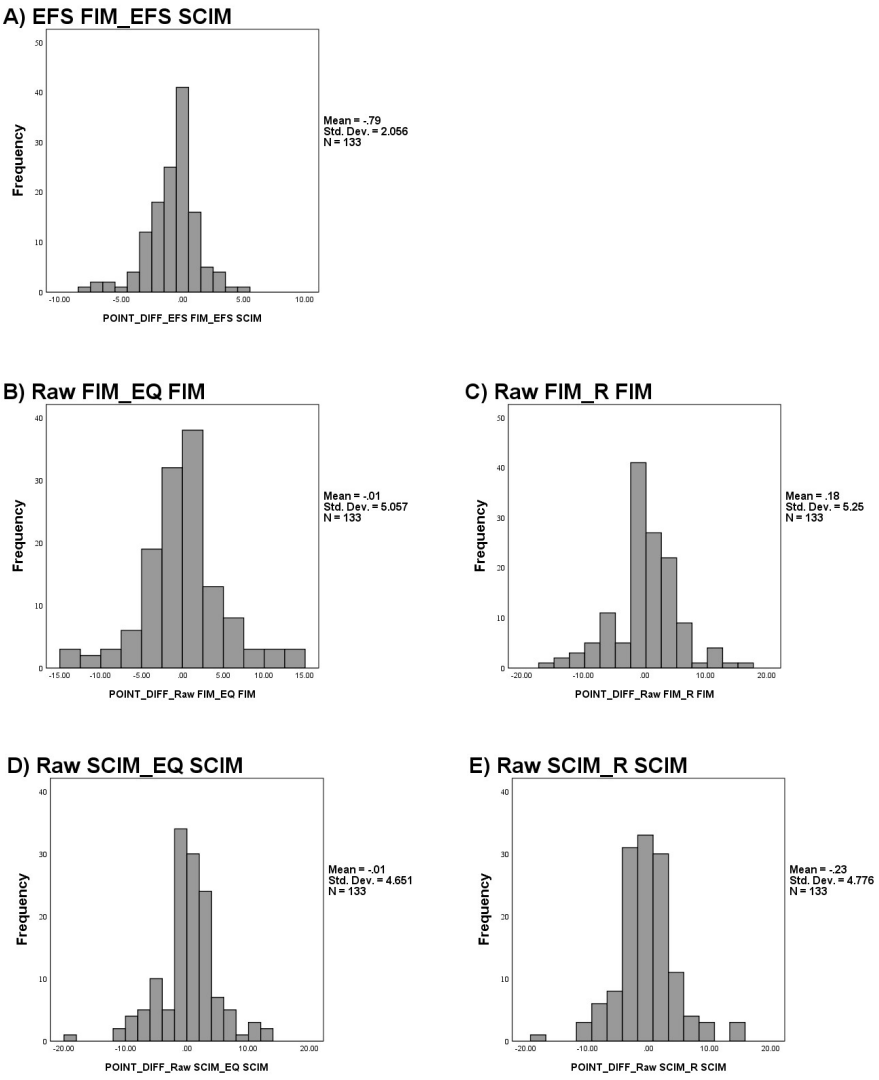
Plot of the frequency of point differences > 0.5 standard deviation from the mean in A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 26: Distribution of point differences SWISS datset



Plot of the frequency of point differences > 0.5 standard deviation from the mean in A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 27: Distribution of point differences Anderson dataset



Plot of the frequency of point differences > 0.5 standard deviation from the mean in A) expert panel FIM/SCIM III (EFS FIM EFS SCIM), B) raw FIM and equipercentile (EQ) FIM, C) raw FIM and Rasch (R) FIM, D) raw SCIM and equipercentile (EQ SCIM), E) raw SCIM and Rasch (R) SCIM

Figure 28: Distribution of point differences RHSCIR dataset

4.5.5 Sub-group invariance

Sub-group invariance was assessed for gender, age, neurological level of injury (para versus tetraplegia) and severity (motor complete vs. incomplete), looking for differences between expert panel FIM and SCIM, raw and crosswalked scores for equipercentile and Rasch FIM and Rasch SCIM III (Table 21). Sub-group invariance in the SWISS dataset was not assessed as demographic data were not available. The pre-set criteria was a difference of less than or equal to 0.8 although prior studies have used either 0.08[81] or 0.11[75].

Across all 50 comparisons, 31/50 were ≤ 0.08, 36/50 were ≤ 0.11, and 48/50 were ≤ 0.20. In looking at patterns in sub-group invariance, differences between raw and crosswalked scores were smallest for age differences in the 31-60 year age category and 61-100 year old group (all <0.08 in both datasets) followed by gender (< 0.8 in RHSCIR, range 0.82- 0.121 in Anderson). Age difference for 0-30 and 31-60, neurological level and severity of injury were similar in terms of exceeding 0.08 (Table 21).

Assessment by method

In examining differences in type of crosswalk, no clear patterns emerge when looking at cut offs ≤ 0.08 (Table 21). When using a cut-off ≤ 0.11, the expert panel method has the most amount of categories below that value (10/50), followed by RFIM and RSCIM (both 7/50). Patterns were similar between neurological level and severity of injury within datasets but not between datasets. For example, in the Anderson data set the methods with the lowest average standardized mean differences was EFS, followed by RFIM, EQFIM, RSCIM and EQ SCIM, which was the same for severity of injury (motor complete vs. incomplete). Findings were similar within RHSCIR for NLI with only one difference in ranking between methods.

Assessment by dataset

The most striking observation is that the Anderson dataset exceeded the criteria of ≥ 0.8 for most comparisons (18/25) across all categories except age 31-60 and 61-100, while RHSCIR only exceeded 0.8 1/25 times (Table 21). However, in both datasets criteria were not greatly exceeded, with the highest RMSD at 0.23 for severity of SCIM raw scores and EQ SCIM (Anderson).

Assessment by outcome measure (FIM/SCIM)

No differences were noted between outcome measures.

Table 21: Subgroup invariance as measured by standardized mean difference for each method and dataset

Sub-group analyzed	SWISS	Anderson	RHSCIR
EFS FIM and EFS SCIM			
Gender	X	0.08	0.03
Age diff 0-30, 31-60	X	**0.09**	**0.09**
Age diff 31-60, 61-100	X	0.08	0.01
NLI	X	0.08	0.03
Severity (A/B vs. C/D)	X	**0.11**	0.01
FIM raw scores and EQ FIM			
Gender	X	**0.12**	0.01
Age diff 0-30, 31-60	X	**0.16**	0.01
Age diff 31-60, 61-100	X	0.04	0.00
NLI	X	**0.19**	0.00
Severity (A/B vs C/D)	X	**0.20**	0.03
FIM raw scores and R FIM			
Gender	X	**0.11**	0.00
Age diff 0-30, 31-60	X	**0.15**	0.01
Age diff 31-60, 61-100	X	0.05	0.01
NLI	X	**0.13**	0.03
Severity (A/B vs C/D)	X	**0.16**	0.03
SCIM raw scores and EQ SCIM			
Gender	X	**0.11**	0.02
Age diff 0-30, 31-60	X	**0.16**	0.01
Age diff 31-60, 61-100	X	0.04	0.01
NLI	X	**0.22**	0.03
Severity (A/B vs C/D)	X	**0.23**	0.02
SCIM raw scores and R SCIM			
Gender	X	**0.11**	0.01
Age diff 0-30, 31-60	X	**0.17**	0.01
Age diff 31-60, 61-100	X	0.03	0.00
NLI	X	**0.19**	0.01
Severity (A/B vs C/D)	X	**0.18**	0.02

Based on available data

EFS= expert panel FIM/SCIM III, EQ = equipercentile, R=Rasch, Age diff= age difference, NLI= neurological level of injury, A= American Spinal Injury Association (ASIA) impairment scale A, B= ASIA impairment scale B, C= ASIA impairment scale C, D= ASIA impairment scale D

Subgroup invariance is determined by the standardized mean difference which is the same formula as and is interpreted in the same way as Cohen's effect size (see Appendix B)

Items in bold exceed 0.08

4.5.6 Additional Analyses

Outliers for all methods (differences between EFS FIM and EFS SCIM, raw FIM and EQ FIM, RFIM, and raw SCIM and EQ SCIM, R SCIM) were examined, based on differences in scores greater than two standard deviations. Table 22 illustrates the percentage of data points greater than two and three standard deviations difference for comparisons for all methods. The percentage of outliers is similar between datasets, with the SWISS dataset demonstrating a slightly lower percentage of outliers.

Table 22: Percentage of outliers for each dataset and method

Number of SDs from the mean	SWISS	Anderson	RHSCIR
> 2 SDs (percent)	10.4%	12.0%	13.0%
> 3 SDS (percent)	1.2%	5.0%	4.5%
> 4 SDS (percent)	0%	0%	0%

SD= standard deviation

The injury characteristics and demographics of data points with outliers were examined for Anderson and RHSCIR (Table 23), where these data were available. Some data points exceeded three standard deviations for multiple comparisons, while most that were two standard deviations different exceeded this figure in multiple comparisons. As there are very few data points exceeding three standard deviations, all analysis presented are for values exceeding two standard deviations (which includes three standard deviations).

The majority of participants that exceeded a two standard deviation difference across all comparisons were motor incomplete, with an average of 80% across both datasets (Anderson - 93%, RHSCIR – 71%). Of those that were motor incomplete, 84% were AIS D (Anderson-77%, RHSCIR – 91%). On average across datasets, 42% of those had paraplegia, 58% tetraplegia. In looking at the average percent of admission vs. discharge, 42% of exams were obtained on admission, 58% at discharge. The

average days between FIM and SCIM III in both datasets combined exams was 0.45, with a mode of 1 and median of -1.

Table 23: Summary characteristics of outliersfor any comparison

	Anderson	**RHSCIR**	**Combined**
Severity (percent, n)			
Motor complete	7% (1/14)	29% (5/17)	19% (6/31)
AIS A (% of all outliers, % of motor complete outliers)	0% 	24% (4/17), 80% (4/5)	13% (4/31), 67% (4/6)
AIS B (% of all outliers, % of motor complete outliers)	7% (1/14), 100% (1/1)	6% (1/17), 20% (1/5)	6% (2/31) , 33% (2/6)
Motor incomplete	93% (13/14)	71% (12/17)	80% (25/31)
AIS C (% of all outliers, % of motor incomplete outliers)	21% (3/14), 23% (3/13)	6% (1/17), 8% (1/12)	13% (4/31), 16% (4/25)
AIS D (% of all outliers, % of motor incomplete outliers)	71% (10/14), 77% (10/13)	65% (11/17), 91% (11/12)	68% (21/31), 84% (21/25)
Level of injury (percent, n)			
Paraplegia/ Tetraplegia	50% (7/14), 50% (7/14)	35% (6/17), 65% (11/17)	42% (13/31), 58% (18/31)
Paraplegia AIS A	0% (0/14)	12% (2/17)	6% (2/31)
Paraplegia AIS B	7.1% (1/14)	0% (0/17)	3% (1/31)
Paraplegia AIS C	14% (2/14)	0% (0/17)	6% (2/31)
Paraplegia AIS D	36% (5/14)	34% (4/17)	29% (9/31)
Tetraplegia AIS A	0% (0/14)	12% (2/17)	6% (2/31)
Tetraplegia AIS B	0% (0/14)	6% (1/17)	3% (1/31)
Tetraplegia AIS C	7% (1/14)	6% (1/17)	6% (2/31)
Tetraplegia AIS D	43% (6/14)	41% (7/17)	48% (15/31)
Time exam obtained (percent,n)			
Admit vs. discharge (d/c)	admit = 36% (5/14) discharge= 64% (9/14)	admit= 47% (8/17), discharge=53% (9/17)	admit= 42% (13/31), discharge= 58% (18/31)
Days between exams			
Mean days between exams	0.28	1.06	-0.45
Mode days between exams	1	-1	1
Median days between exams	1	-1	-1

A = American Spinal Injury Association (ASIA) impairment scale A, B= ASIA impairment scale B, C= ASIA impairment scale C, D= ASIA impairment scale D

Specific differences in items and categories of differences also were examined for the EFS FIM and SCIM scores, for individual datasets that varied by two SDs in any comparison. Combined data for Anderson and RHSCIR demonstrated that of the 122 times where a difference in EFS occurred for any item, 85% of the time, the EFS FIM score was lower than the EFS SCIM score. One possibility was that FIM was administered earlier than SCIM and that the differences were due to true change. However, in examining the days between FIM and SCIM III and which exam occurred first, no clear pattern emerged. Generally, FIM scores were lower when FIM was administered first or second longitudinally.

In looking at patterns in differences for individual EFS scores where scoring options are 0-3 (0-total assistance, 1-partial assistance,2-independent except with equipment, 3-independent without equipment) items are (eat, bath, dress groom , use of toilet, stairs), the majority of differences are found with an EFS FIM of partial assistance and a SCIM III of independent with equipment or set-up and FIM independent with equipment or set-up with a SCIM III of independent without equipment or set-up. In looking at differences for the transfer items (transfer bed, transfer tub/toilet) where only three scoring options are available (0-total assistance, 1-partial assistance, 2-independent without equipment), the most common difference was between a FIM of Partial Assistance and a SCIM of independent without equipment or set-up. The walk/wheelchair EFS item consisted of 5 possible scores (0-total assistance walk or wheelchair, 1-uses wheelchair with or without assistance, 2-walks with assistance or supervision with or without equipment, 3-walks independently with equipment, 4- walks independently without equipment). Most of the differences in this category appear related to the mode of locomotion. For example, an individual who uses a wheelchair (EFS FIM score of 1) might have an EFS SCIM score of walks with equipment or vice versa.

The most common items with score differences were (in order of highest to lowest) eating, dressing upper body, grooming, dressing lower body, bathing/use of toilet/walk/wheelchair, stairs, bed

transfers/tub/toilet transfers. The most common discrepancies (at least four or more individual data points with a discrepancy in a given category) for each item are summarized below (Table 24).

Table 24: Most common discrepancies between expert panel FIM and expert panel SCIM for individual items

	Most common discrepancy	Second most common discrepancy	Third most common discrepancy
Eating	EFS FIM of independent with equipment or set-up and EFS SCIM of independent without equipment or set-up	EFS FIM of partial assistance and EFS SCIM of independent with equipment or set-up	EFS FIM of partial assistance and EFS SCIM of independent without equipment.
Bathing	EFS FIM partial assistance and EFS SCIM independent with equipment or set-up		
Dressing	EFSFIM partial assistance and EFS SCIM independent with equipment or set-up	EFS FIM of partial assistance and EFS SCIM of independent with equipment or set-up	
Grooming	EFS FIM partial assistance and EFS SCIM independent without equipment	EFS FIM of independent with equipment or set-up and independent without equipment or set-up	
Use of toilet	EFS FIM of independent with equipment or set-up and independent without equipment or set-up	EFS FIM of partial assistance and EFS SCIM of independent with equipment or set-up	
Transfer items	EFS FIM partial assistance and EFS SCIM independent without equipment		
Stairs	EFS FIM of partial assistance and EFS SCIM of independent with equipment or set-up		

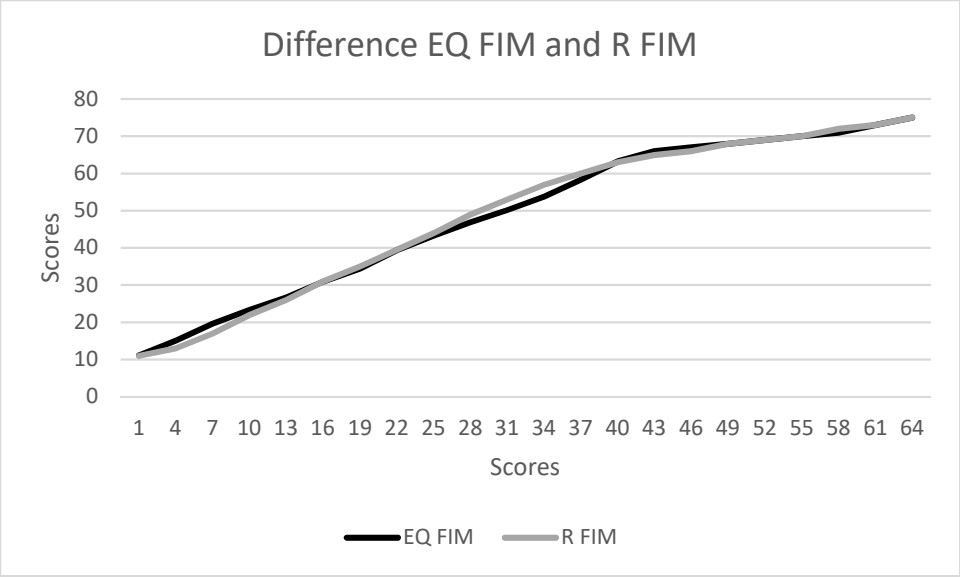
EFS= expert panel FIM/SCIM III

4.5.7 Summary Comparisons

Summary comparison of methods

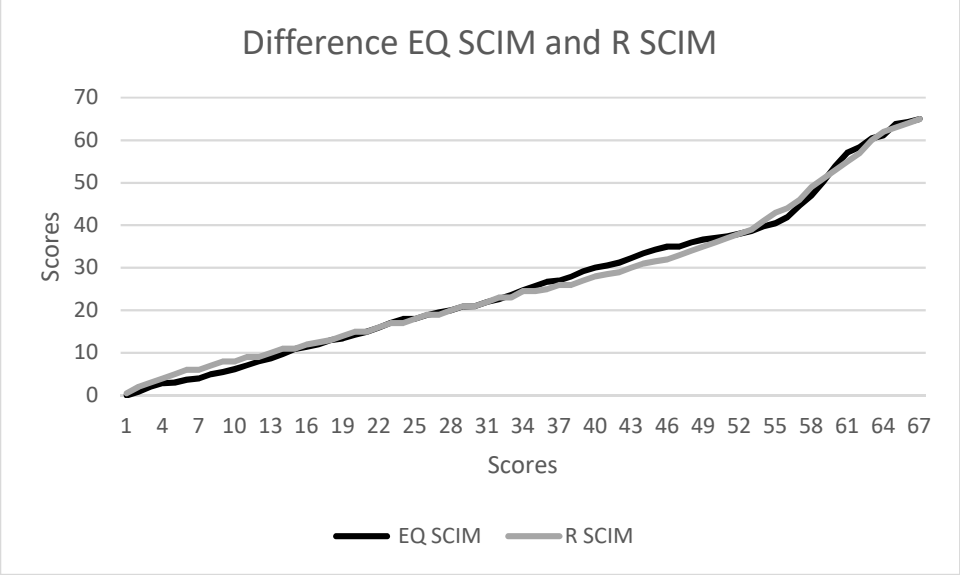
The correlation coefficients across all methods did not demonstrate any appreciable differences (expert panel= 0.911 – 0.972, EQ SCIM= 0.901 – 0.969, EQFIM=0.918 – 0.971, RSCIM = 0.897 – 0.966, RFIM= 0.917 – 0.971). Crudely, one can summarize how many of each primary criteria and additional criteria were met using the correlation coefficient, score distributions (if means or standard deviations exceeded criteria), Bland Altman plots (if one method appears more biased in its distribution), Cohen's effect size, point differences, and the five categories for sub-group invariance. EFS meets criteria 56% (14/25), EQ FIM 84% (21/25), EQ SCIM 76% (19/25), R FIM= 76% (19/25) and RSCIM 80% (20/25), suggesting that the expert panel method is less strong when additional criteria are considered. The equipercentile and Rasch methods appear quite similar.

Figure 29 and Figure 30 present a graphic comparison of equipercentile and Rasch crosswalk tables. Although two conversions were created and are displayed for equipercentile and Rasch, the differences are related to rounding, thus the graphs are inverse and similar. The majority of differences are noted in the lower and middle of the score ranges, with the largest difference of 4 points in the SCIM raw score to EQ FIM and R FIM conversions. Most differences are 0-1 point differences.



Plot of scores for equipercentile FIM (EQ FIM) and Rasch (R FIM)

Figure 29: Graphical display of differences between equipercentile FIM and Rasch FIM crosswalk



Plot of scores for equipercentile SCIM (EQ SCIM) and Rasch SCIM (R SCIM)

Figure 30: Graphical display of differences between equipercentile SCIM and Rasch SCIM crosswalk

Summary comparison of databases

Correlation coefficients were (on average) lowest in the SWISS dataset (0.910), and highest for the RHSCIR dataset (0.970), with Anderson falling in the middle (0.924).

As noted, in the Anderson dataset, a number of secondary criteria were not met. Crudely, in examining how many criteria were met in each dataset, SWISS met 92% (23/25), Anderson 46% (23/50) and RHSCIR 94% (47/50). As SWISS did not contain demographic data such that sub-group invariance could not be computed, this high percentage of criteria that were met may be inflated artificially as the denominator is lower.

Summary comparison by outcome measure (FIM/SCIM)

In grossly examining primary and secondary criteria there were no meaningful differences between FIM and SCIM III crosswalks. As discussed above, FIM scores appear to be lower by comparison than SCIM III scores.

CHAPTER

V. DISCUSSION

Three crosswalks for FIM and SCIM III voluntary motor scores were developed and validated using three different methods. All crosswalks met the primary criteria correlation coefficients > 0.866 (to reduce uncertainty by 50%) although there was some variability in assessing secondary criteria.

5.1 Assessment of dimensionality

A key factor in determining if two instruments should be linked is that they are unidimensional, meaning they assess a single construct. In this case, the underlying construct is voluntary motor function, as items reflecting autonomic dysfunction (bowel, bladder, respiration) represents a different construct. Initial correlations between raw FIM/SCIM III scores were strong across all datasets (range 0.898 – 0.961) implying a single construct. As part of the expert panel method, all items in both instruments were reviewed. Reviewers did not comment that items clearly separated into two dimensions. More objective methods were used to assess dimensionality, exploratory factor analysis and Rasch principal components analysis of residuals. Although there is evidence that two or possibly three underlying dimensions exist (relative activities reflecting predominant use of upper extremity vs. lower extremity function), the dimensions were not strong or clear enough to warrant a multi-dimensional approach. Additionally, not all items clearly associated with one dimension or another. Therefore, a single crosswalk for each method was created, reflecting a unidimensional construct of voluntary motor function.

This is in contrast to the approach of Hong et al.[20] who developed three FIM and Korean Modified Barthel Index (K-MBI) crosswalks for self-care, mobility and involuntary motor function items,

in a mixed diagnostic group. In the Hong study items clearly loaded on self-care, mobility and involuntary movement, based on exploratory factor analysis. Although it is unclear what approach within exploratory factor analysis was used (orthogonal or oblique rotation), which impacts whether results are presented as correlations coefficients or regression coefficients, there are some interesting differences between the Hong et al.[20] and current FIM/SCIM III analyses. Items such as FIM lower body dressing, which clearly loaded on one factor in their analysis (0.851) were split between two factors in the FIM/SCIM III analysis (factor 1= 0.524, factor 2= 0.387). A similar situation occurred with the FIM/SCIM III assessment where FIM use of toilet, bed transfer, toilet transfer, tub transfer items, did not clearly load on one factor or another. Interestingly, some items which clearly loaded on the first factor in the FIM/SCIM III analysis, were not as strong in the FIM/K-MBI analysis, such as feeding (0.543 on first factor in Hong, 1.005 in this analysis). The variance explained was similar between the FIM/KMBI and FIM/SCIM III analyses with approximately 60% loading on first factor and approximately 10% on the second factor. Differences between analyses may be due to the use of a mixed diagnostic group in the Hong study and a SCI specific group in this analysis.

The assumption of a unidimensional construct in this analysis allows the use of a single crosswalk vs. a different crosswalk for each identified dimension. A single crosswalk is easier to use as one only needs to apply the crosswalk to a single score. However, as items representing autonomic function were not included in the crosswalk, future use will require access to voluntary motor item scores (the motor category in FIM) or to individual item scores. In this analysis, use of toilet was considered a voluntary motor function item (vs. autonomic function) as it relates to perineal hygiene, adjustment of clothing, and napkins/diapers vs. actual bowel or bladder function. However, in SCIM III use of toilet is part of respiratory and sphincter management, thus individual items scores are required to use the crosswalk.

The assumption of a unidimensional construct confers bias. Using crosswalks for each potential factor may have resulted in stronger crosswalks, with higher correlations and less error.

5.2 Findings by Method

As noted, all methods met the primary criteria of correlation coefficients greater than 0.866. The expert panel method resulted in strong correlations across all three datasets, despite the loss of data which occurred by collapsing categories and scores and items which had no matching FIM item but contribute to the SCIM III total score. Despite the fact that this method is not statistically based, but is based on expert panel input, which can be subjective, correlations are strong. One possibility is that the items that were not used (SCIM mobility in bed and action to prevent bed sores, mobility indoors and outdoors, ground to wheelchair transfer, wheelchair to car transfer) are less reliable or demonstrate item misfit or discordant score thresholds. In looking at reliability for these items, most had high reliability (percent agreement between raters - 77.1% to 96.2%) in the Itzkovich et al.[32] reliability study with somewhat lower reliability in the Anderson et al. study[1] (percent agreement between raters – 68% to94%).

Although Catz et al.[46] completed a Rasch analysis by SCIM III sub-groups, which produces different results than co-calibrating all items, mobility outdoors exhibited slightly high infit and outfit values (infit= 1.51, outfit= 1.56) using the criteria of 0.8-1.4 used by Catz. High infit values were also found for score categories for mobility indoors, moderate distances and outdoors (score of 5 = walks with crutches/canes) and ground to wheelchair transfers (score of 0= total assistance). Disordered score thresholds were identified for mobility for all distances. In the current analysis, there was evidence to suggest that mobility items for all distances disordered score thresholds. The fact that only one mobility item was used (moderate distance) in the expert panel method and ground to wheelchair transfers was

not used, may have contributed to the high correlation coefficient for this method, as items with disordered score thresholds were not utilized. Other items which demonstrated misfit in the Catz et al.[46] analysis (feeding) were retained.

In comparing with Reed et al.[66], based on expert feedback and initial assessments of Rasch score threshold disorder, scoring was four categories in most cases: complete dependence on assistance from others, partial dependence on assistance from others, independence with devices and independence without devices, while grooming was reduced to three categories: complete dependence on assistance from others, partial dependence on assistance from others, independence with or without devices. The categories of scoring options used by Reed et al.[66] are similar to what was used in the expert panel method where eating, bathing, dressing, and use of toilet had four possible scores with the same description. In the EFS FIM/SCIM, grooming and stairs had four scoring options while Reed et al.[66] (personal communication John Steeves, February 2020) reduced scoring options to three. EFS FIM/SCIM mobility consisted of fiver options in order to reflect mode of locomotion in FIM, while Reed et al.[66] (personal communication John Steeves, February 2020) reduced this item to four scoring options.

Thus, the elimination of indoor and outdoor mobility items, and the reduction of scoring options may have led to a high correlation coefficient as the data that was "lost" did not add (and perhaps even detracted) from the model.

In looking at secondary criteria, the expert panel method did not perform as well as other methods. For example, EFS FIM and EFS SCIM means were all significantly different across all three datasets. In addition, Cohen's effect sizes ranged from medium (0.38) to large (0.65) across all three data sets for this method while the only point differences that did not meet the pre-established criteria of 75% within ½ standard deviation was in the EFS method (albeit 74% which was only one percentage point lower then defined criteria). These differences may be a result of the fact that items and scores

were compressed such that any differences between FIM and SCIM III were magnified, relative to other methods where all items and scores are preserved.

Given the loss of data and weaker assessment of secondary criteria for the expert panel method, consideration is now given to the equipercentile and Rasch methods, which are both statistically based and resulted in high correlations. Equipercentile demonstrated one mean score difference (EQ SCIM) in the Anderson dataset, one effect size slightly higher than >0.20 (0.21 – EQ SCIM in Anderson dataset), with a number of areas in which subgroup invariance exceeded criteria (EQ FIM and EQ SCIM in Anderson dataset).

The Rasch crosswalk method resulted in significantly different means for both FIM and SCIM III raw core to crosswalk conversions as well as standard deviations with a greater than one-unit difference. Cohen's effect size slightly exceeded 0.20 for both raw score to SCIM III (0.24) and raw score to FIM (0.28) crosswalk scores. Sub-group invariance exceeded criteria of 0.08 for raw score to RSCIM and raw score RFIM in most categories except age difference of 31-60 and 61-100. In all cases, where secondary analyses exceeded pre-determined cut-offs, this occurred in the Anderson dataset which will be discussed below in relation to differences between datasets below.

Although, as identified above, pre-established criteria were exceeded for some of the secondary crosswalk assessed, in most cases where values were exceeded, the difference was small. Using the pre-established criteria of correlation coefficients and secondary analysis for crosswalk validation and assessment, neither the Rasch nor equipercentile crosswalk appear superior. This is confirmed in Figure 29 and Figure 30, where the equipercentile and Rasch crosswalks are very similar. However, the equipercentile crosswalk aligns original, ordinal measures (based on classical test theory), while the Rasch crosswalk is based on co-calibration of linearized measures on a logit scale. Thus, more advanced parametric (vs. non parametric) statistics can be applied. Additionally, logits are provided for both

individual items and total scores, allowing comparisons to other interval or linearized measures. Thus, use of the Rasch based FIM/SCIM III crosswalk is recommended.

5.3 Findings by dataset

The secondary crosswalk criteria from the Anderson validation dataset consistently did not meet pre-set criteria, to a noticeably greater extent than the SWISS and RHSCIR datasets. There are a number of possible explanations.

The Anderson dataset was collected as part of a SCIM III reliability study. As the intent was to use SCIM III as presented by the developer, which does not include explanations or training, no SCIM III training was conducted and further SCIM III is not routinely collected in the US, so the examiners collectively had less experience with SCIM III than those in the SWISS and RHSCIR datasets. SCIM III reliability was strong based on the study findings[1], although not as strong as a prior study by the instrument developers.[32] This was attributed to the lack of a guidance manual on the use of SCIM III.[1] It is possible, but not likely that both examiners for interrater reliability were reliable (similar findings), but both inaccurate.

Another contributing factor may be the percentage of individuals in the Anderson dataset with an AIS D injury severity. Although not statistically significant, there was a higher percentage of AIS D in the Anderson data set relative to RHSCIR (49% vs. 39%). As discussed below under analysis by outcome measure, in looking at outlier data points, the majority of outliers were individuals with AIS D (%). Although there are no publications which specifically consider reliability by AIS grade, Segal et al.[35] (n=57) found that reliability for FIM was lowest in individuals with incomplete tetraplegia (0.49, n=17) relative to other impairment groups (complete tetraplegia – 0.87, n=14, complete paraplegia - 0.74, n=13,), incomplete paraplegia - 0.85, n=9). No data for reliability for SCIM III in in relation to impairment (level or severity of injury) has been published.

The total length of stay and time from admission to final FIM and SCIM III exam was significantly longer in RHSCIR than Anderson (139 days vs. 76 days). The days between FIM and SCIM at discharge averaged two days in Anderson and zero days in RHSCIR, and due to weighted selections of discharge over admission data in Anderson (Anderson DC=81%, RHSCIR DC= 44%), there is a significantly higher percentage of discharge than admission data in Anderson vs. RHSCIR. Intuitively, the longer the length of stay, the less change will occur over a short period of time. Given the high percentage of discharge time points used for the analysis, the shorter length of stay and higher number of days between administration of FIM vs. SCIM III, there may have been true differences in function between the time FIM was collected and the time SCIM III was collected in Anderson, impacting the comparison of raw to converted FIM and SCIM III scores. The Anderson dataset had a significantly lower percentage of individuals with tetraplegia (47%), relative to RHSCIR (69%), based on available data (49% of data in Anderson, 77%). These figures both differ from the current US national average of 59.5% for tetraplegia[4] (although during the time frame the Anderson data was collected the percentage was lower - 51.3% to 53.4%), but this may be due to the large percent of missing data. It is likely a combination the percent of individuals with AIS D injuries and the likelihood that AIS D exams are less reliable, that contributed to the consistent finding that the Anderson dataset did not meet criteria in secondary analyses.

Additional differences noted between datasets include a significant difference between FIM mode, when examining all three datasets. This difference is due to comparisons between the SWISS dataset relative to RHSCIR and Anderson. The SWISS data set had a much lower percentage of individuals walking at the time the data were collected (9%), than Anderson (27%) and RHSCIR (31%). This may reflect differences in the way the data were collected and systems of care. SWISS data were collected at specific cross sectional time points, unrelated to admission and discharge. The length of stay in Switzerland is much longer than either Canada or the US (140, 73, 34 days respectively). In the US the

focus of inpatient rehabilitation is on compensation in order to discharge the patient home with skills required for safety (e.g. transfers) where locomotor training will largely occur on an outpatient basis. In Switzerland, due to the long length of stay, the bulk of locomotor training occurs on an inpatient basis. As we do not know the time point for data collection in the SWISS dataset relative to discharge, it may be that the time points on the randomized data fell earlier in the rehabilitation stay. Alternatively, UDS guidance for FIM requires that the predominant locomotion modifier at discharge is "back coded" to admission. Thus, if an individual uses a wheelchair at admission and the FIM wheelchair modifier is assigned, if an individual's primary mode of locomotion is walking at discharge, the admission mode will be recoded to walk. As FIM data collected in SWISS was not necessarily related to admission and discharge, admission codes likely were not adjusted. As there is no significant difference in FIM and SCIM III overall scores between databases (Swiss scores fall between Anderson and RHSCIR) SWISS participants did not have lower function than participants in the other datasets, thus the low percent of walk mode is likely due to the lack of adjustment in the walk vs. wheelchair coding.

SCIM III motor scores at discharge were significantly higher between Anderson and RHSCIR (data unavailable for SWISS), which is not surprising considering the longer lengths of stay in Canada. Although not significantly different, FIM scores were also higher at DC for RHSCIR.

Although the percentage of outliers/invalid data points in the development dataset was larger than what is expected in a normal distribution (0.3% >3 standard deviations from the mean, 6%-person misfit in Rasch), the percentage is low and unlikely to impact crosswalk accuracy. The decision to set a relatively high cut-off of 10% was made in order to use real world data vs. "cleaned" data as outliers/invalid data points can be expected when the crosswalk is used in research or clinical settings. Differences in outliers/invalid data in different settings were noted even within this project. Using the criteria of greater than three standard deviations from the mean, the SWISS database, which was

created solely for research purposes, had 1.2% outliers, while the other research database (Anderson) had 5.0% outliers, and the registry database (RHSCIR) 4.5% outliers. Anderson and RHSCIR are probably the closest to real world data. Although Anderson data resulted from a research study, no specific training was provided. Although specific training is provided in RHSCIR, the rigor of data collection and cleaning typically is lower in registries relative to research databases.

Both the Anderson and RHSCIR datasets had a low percentage of usable data (31% and 24% respectively). However, these figures do not necessarily represent data quality. The Anderson data set was derived from a SCIM III reliability study, with a comparison to FIM for responsiveness, where mode was not used. Only data within the dataset that contained mode were used for the FIM/SCIM III crosswalk, as mode was required for the expert panel method. Had mode not been required, the dataset would have been much larger. In the RHSCIR dataset, SCIM III is collected by observation or clinician questionnaire. In the FIM/SCIM III crosswalk, only SCIM III collected by observation was used for consistency with other datasets.

5.4 Findings by measure

A number of interesting findings were noted relevant to FIM and SCIM III as part of this work including challenges with application of measures in sub-populations, construct similarities between measures and similar item difficulties.

In examining outliers in Anderson and RHSCIR, the majority were AIS D (68% of all outliers) and of these, 38% were individuals with paraplegia, 62% with tetraplegia. Similar to the discussion above about the findings in the Anderson dataset, outliers could be due to true change, issues with exam reliability, coding errors or inconsistencies in the way FIM and SCIM III captures these individuals. If the difference was due to true change between the FIM and SCIM III exams, one might expect that a larger percentage of these exams occurred at admission, with a large number of days between exams. True

change over a short period of time is more likely earlier post injury than later in recovery, particularly in RHSCIR with a significantly longer length stay than Anderson. Individuals with an AIS D injury are more likely to experience true change as less severe injuries have less recovery over time. However, in examining the percentage of outliers at admission versus discharge, the percentage of admission outliers (42%) was lower than discharge outliers (58%) and the mean days between exams was only 0.45 (mode=1, median=1). So, it is not likely that true change is what caused the outliers, but perhaps the measures are less reliable and more variable due to wide spectrum of function for individuals with an AIS D injury severity. Reliability in the Anderson et al.[1] SCIM III reliability study was noted to be lower than the Izkovitch et al.[32] reliability study, but no studies have reported on reliability issues in relation to severity of SCI.

As noted in outlier analysis, when comparing EFS FIM to EFS SCIM, EFS FIM scores were lower. One possibility is that for most items EFS collapsed multiple FIM categories (maximal assistance, moderate assistance, minimal contact assistance and for some items supervision or set-up), whereas SCIM III often has one category for partial assistance, or sometimes two. The collapsing of FIM categories may have resulted in lower EFS FIM scores. Anderson et al.[1] recommended expansion of the partial assistance category as an individual who only requires supervision is rated in the same way as someone requiring maximal assistance. The recommendation was to split this category into “requires 50% or more assistance”, “requires 50% or less assistance”, “requires supervision or device only”. This would equate to collapsing FIM categories of 1-2, 3-4 and 5-6. Given the score disorder noted by Catz et al.[46] and Reed et al.[66] such changes would need to be considered carefully, validated and subjected to Rasch analysis.

Given the high correlation coefficients, similarities in most items in terms of item difficulty as assessed by logits, factor and Rasch analysis of dimensionality, FIM and SCIM III voluntary motor items

appear to assess a similar underlying construct between measures. Although FIM was intended to capture burden of care while the SCIM developers focused on effectiveness of rehabilitation training and items specifically reflecting what is important to individuals living with spinal cord injury, the distinction between the two is not strong. The similarities may be less consistent when bowel and bladder items are included as well as respiratory function which is not present in FIM. That being said, SCIM III has shown greater responsiveness to change for some items (respiration and sphincter management in Itzkovich et al.[32] and Anderson et al.[1], mobility indoors and outdoors in Itzkovich et al.[32]), than FIM in the SCI population and is currently collected throughout Europe. Canada collects both FIM and SCIM III, while Australia primarily collects FIM. The United States is now collecting the follow-up measure to FIM, the CARE Tool (although FIM is still collected in facilities designated as rehabilitation facilities) and SCIM is not collected routinely. A SCIM III manual currently is being developed using a modified Delphi process, to standardize SCIM III by observation (personal communication MJ Mulcahey, February 2020), which may improve reliability and ease of use.

As illustrated in Table 24, discrepancies were noted in EFS scores for FIM and SCIM III in certain items more than others (eating with most discrepancies, stairs with least), and for certain scores more than others (differences in partial assistance, independent with equipment and without equipment). Although the differences were highlighted in the EFS crosswalk, these differences relate back to original FIM and SCIM III data, thus this information can be used to improve training in problematic items and scores.

5.5 Additional findings

5.5.1 Individual vs. group level use of the crosswalk

In assessing the utility of the crosswalk for use at the group vs. individual level, the findings support crosswalk use at the individual participant level as well as at the group level. Different

researchers use differing criteria for this differentiation. Most researchers concur that correlation coefficients should exceed 0.866 for individual level analyses. Point differences also are commonly assessed for use at the individual level. In this study, point differences were within the pre-set criteria of 75% within 0.5 standard deviations. The average for the expert panel method was 83.1%, while equipercentile and Rasch averages were 89% as were FIM and SCIM III point differences. All of these, especially equipercentile and Rasch are quite high, supporting the use of the crosswalk at the individual level.

5.5.2 Comparison with prior FIM and SCIM III Rasch analyses

A benefit of Rasch analysis is that items are identified that do not fit the primary construct, exhibit misfit, disordered scoring, or are redundant as identified by item misfit. As noted, a number of items demonstrated infit and or outfit values that exceeded the threshold of 0.5-1.7[110], recommended for measures assessed by clinical observation. In prior Rasch analyses of SCIM III, the authors used a misfit range of 0.8 – 1.4[46], while for a Rasch analysis of FIM used 0.7-1.3.[115] There are a number of similarities and differences between the current analysis and prior Rasch analyses.

In the FIM Rasch analysis conducted by Linacre et al.[115], only data with the functional modifier for walk was used and this item "walk" did not demonstrate item misfit. In the current FIM/SCIM III analysis, FIM walk/wheelchair demonstrated high infit and outfit values, which may indicate it does not fit the unidimensional construct. However, it is likely this value reflects the fact that mode of locomotion (walk vs. wheelchair) was not included in the model. The impact of this is that an individual with a low score, could have a low score for wheelchair locomotion or walking, which may not be consistent across other items. For example, an individual with a low score due to locomotion may have a high score in other items reflecting lower extremity function, whereas an individual with a low score due to wheelchair use will not likely exhibit high scores across lower extremity behaviors. So, this item may in

fact fit the unidimensional model, but the functional modifier needs to be considered in the model. For voluntary motor items, Linacre et al.[115] also found item misfit for eating, and stairs, which was not found in this analysis, possibly due to differences in the model. For this analysis, a group rating scale was used, where the partial credit model may have different findings.

As discussed above, Catz et al.[46] identified item misfit in voluntary motor items for feeding, mobility outdoors and stair management, with disordered scoring thresholds for use of toilet, mobility for all distances and stairs. In this FIM/SCIM III analysis, the only SCIM III item with item misfit was bed mobility. Items with disordered score thresholds were SCIM mobility items for all distances, indoors, moderate distance and outdoors, which also demonstrated evidence of not fitting the unidimensional model. For the SCIM mobility items, scoring is across eight categories, with an area of potential score disorder in scores 3-8 where use of different aids is reflected: Supervision (3), walking frame or crutches-swing (4), crutches or two canes (5), one cane (6), leg orthosis only (6). Additionally, the use of devices may reflect personal choice more than ability in some individuals, which may contribute to disordered score thresholds.

5.6 Study Limitations

Common person equating using retrospective data were used for the analysis. The disadvantage of using retrospective data is that it was collected for a different purpose, thus data may be less accurate than if data are collected and personnel trained for a specific purpose. Conversely, the retrospective data used for this analysis reflect real world data, in which training and accuracy of data collection may be less, but more applicable to future uses of the crosswalk.

Although a unidimensional approach was pursued, given that some elements of both exploratory factor and Rasch analysis could have supported multiple constructs, creating separate crosswalks for items clearly focusing on upper or lower extremity function may have resulted in stronger

crosswalks. Conversely, using a single crosswalk for total scores, vs. sub-scores is more straightforward for the user.

A related potential limitation is that a core aspect of Rasch analysis is the ability to remove items with misfit and rescore disordered score thresholds so the data can fit the model. Although this approach would likely improve the model and thus the crosswalk, this approach was not used as this would result in loss of data and would make real world application of the crosswalk more challenging and less straightforward than using total scores.

Both a strength and a challenge of the data used in this analysis is the fact that three databases, created for different purposes, from three different countries with different systems of care were used in the analysis. The strength of this approach is that the findings are more generalizable. Challenges included loss of data due to missing and out of range values, missing data on FIM modifiers and a mixture of clinical observation and clinician questionnaire for SCIM III, limiting the data set for consistency using only SCIM III data collected by questionnaire. An additional limitation is that FIM data in RHSCIR are obtained via linkage with hospital administrative data which were collected via the clinician questioning the participant. Additionally, FIM and SCIM III data were not always collected on the same day such that function may have changed over the time frame both measures were collected. This was, in part, accounted for by limiting the time frame between exams to seven days, to minimize the likelihood of true change. Conversely, where FIM and SCIM III were collected on the same day, fatigue and learning may have impacted the findings. In the Anderson dataset the collection of a single SCIM III assessment sometimes occurred over multiple days. A significant challenge in the SWISS development dataset is that no demographic data were available and the time of data collection relative to injury was not available.

5.7 Future Directions

The intent of creating a crosswalk is to enable comparisons across databases using one or the other measure. An effort currently is underway to compare neurological data from three SCI registries. The FIM/SCIM III crosswalk will pave the way to extending this work to examine similarities and differences in neurological recovery. Charlifue et al.[58] provided an overview of existing databases to "facilitate international collaborations and enhance comparability, data pooling, and the ability to generalize findings to a broader population". The creation of the crosswalk is a critical step in facilitating comparisons of databases, in this era of "big data". Next steps involve examination of multiple datasets across different countries and systems of care to look for similarities and differences in patterns of functional recovery.

Additional assessments of FIM and SCIM III crosswalks to include involuntary motor functions are important as this is an area where the instruments may differentiate themselves for individuals living with SCI. As a follow-up to the current work, SCIM III collected by clinician report could be included in the crosswalk to expand generalizability.

In September 2018, CMS and thus Model Systems transitioned to the Care Tool. As FIM is no longer routinely collected in all US centers, a SCIM III Care Tool crosswalk is warranted. Additionally, a validation study of SCIM IV is underway. Proposed changes include scoring for nine transfer items, where items differ based on whether an individual can walk or uses a wheelchair. Unfortunately, there are no proposed changes addressing item misfit or score threshold disorder. Although there is concern in the SCI community about adoption of a new version of SCIM (SCIM IV), if it is validated and adopted, a SCIM IV crosswalk may be needed.

5.8 Conclusion

In summary, the Rasch FIM/SCIM III crosswalk is recommended for use at the group and individual level. Further understanding of the reliability of both FIM and SCIM III in individuals with AIS D (particularly tetraplegia) are warranted. Findings related to discrepancies in scoring between the two measures, can be utilized in training on these assessments. Ideally, item misfit, dimensionality and disordered scoring would be addressed by SCIM III instrument developers. Further crosswalk development should consider the Rasch approach due to the ability to linearize measures and improve understanding of the strengths and challenges of the measures to be linked.

REFERENCES

1. Anderson KD, Acuff ME, Arp BG, et al. United States (US) multi-center study to assess the validity and reliability of the Spinal Cord Independence Measure (SCIM III). *Spinal Cord.* 2011;49(8):880-885.
2. Hall KM, Cohen ME, Wright J, Call M, Werner P. Characteristics of the Functional Independence Measure in traumatic spinal cord injury. *Arch Phys Med Rehabil.* 1999;80(11):1471-1476.
3. Furlan JC, Sakakibara BM, Miller WC, Krassioukov AV. Global incidence and prevalence of traumatic spinal cord injury. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques.* 2013;40(4):456-464.
4. National Spinal Cord Injury Statistical Center, Facts and Figures At a Glance. Birmingham, AL: University of Alabama at Birmingham, 2020. https://www.nscisc.uab.edu/Public/Facts%20and%20Figures%202020.pdf. Accessed April 9, 2020.
5. Steeves JD, Lammertse D, Curt A, et al. Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures. *Spinal Cord.* 2007;45(3):206-221.
6. Lammertse DP, Jones LA, Charlifue SB, et al. Autologous incubated macrophage therapy in acute, complete spinal cord injury: results of the phase 2 randomized controlled multicenter trial. *Spinal Cord.* 2012;50(9):661-671.
7. Spinal Cord Outcomes Partership Endeavor: Clinical Trials. http://scope-sci.org/trials/. Accessed 5 April, 2020.
8. Dorans NJ. Equating, Concordance, and Expectation. *Applied Psychological Measurement.* 2004;28(4):227-246.
9. Dorans NJ, Holland PW. Population invariance and the equatability of tests: basic theory and the linear case. *ETS Research Report Series.* 2000;2000(2):i-35.
10. World Health Organization. How to use the ICF: A practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment. WHO. Published October 2013. Accessed February 21, 2017.
11. Jones LAT, Bryden A, Wheeler TL, et al. Considerations and recommendations for selection and utilization of upper extremity clinical outcome assessments in human spinal cord injury trials. *Spinal Cord.* 2018;56(5):414-425.
12. Biering-Sorensen F, Alai S, Anderson K, et al. Common data elements for spinal cord injury clinical research: a National Institute for Neurological Disorders and Stroke project. *Spinal Cord.* 2015;53(4):265-277.
13. Granger CV. The emerging science of functional assessment: our tool for outcomes analysis. *Arch Phys Med Rehabil.* 1998;79(3):235-240.
14. Msall ME, DiGaudio K, Rogers BT, et al. The Functional Independence Measure for Children (WeeFIM). Conceptual basis and pilot use in children with developmental disabilities. *Clinical pediatrics.* 1994;33(7):421-430.
15. Granger CV, Hamilton BB, Keith RA, Zielezny M, Sherwin FS. Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation.* 1986;1(3):59-74.
16. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil.* 1987;1:6-18.
17. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the functional independence measure. *Arch Phys Med Rehabil.* 1993;74(6):566-573.

18. Stineman MG, Shea JA, Jette A, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. *Arch Phys Med Rehabil.* 1996;77(11):1101-1108.
19. Stineman MG, Jette A, Fiedler R, Granger C. Impairment-specific dimensions within the Functional Independence Measure. *Arch Phys Med Rehabil.* 1997;78(6):636-643.
20. Hong I, Woo HS, Shim S, Li CY, Yoonjeong L, Velozo CA. Equating activities of daily living outcome measures: the Functional Independence Measure and the Korean version of Modified Barthel Index. *Disabil Rehabil.* 2018;40(2):217-224.
21. Rehabilitation Measures Database: Functional Independance Measure. https://www.sralab.org/search?contains=functional%20independence%20measure. Accessed.
22. Marino RJ, Huang M, Knight P, Herbison GJ, Ditunno JF, Jr., Segal M. Assessing selfcare status in quadriplegia: comparison of the quadriplegia index of function (QIF) and the functional independence measure (FIM). *Paraplegia.* 1993;31(4):225-233.
23. Yavuz N, Tezyurek M, Akyuz M. A comparison of two functional tests in quadriplegia: the quadriplegia index of function and the functional independence measure. *Spinal Cord.* 1998;36(12):832-837.
24. Saboe LA, Darrah JM, Pain KS, Guthrie J. Early predictors of functional independence 2 years after spinal cord injury. *Arch Phys Med Rehabil.* 1997;78(6):644-650.
25. Marino RJ, Shea JA, Stineman MG. The Capabilities of Upper Extremity instrument: reliability and validity of a measure of functional limitation in tetraplegia. *Arch Phys Med Rehabil.* 1998;79(12):1512-1521.
26. Morganti B, Scivoletto G, Ditunno P, Ditunno JF, Molinari M. Walking index for spinal cord injury (WISCI): criterion validation. *Spinal Cord.* 2005;43(1):27-33.
27. Ditunno JF, Jr., Barbeau H, Dobkin BH, et al. Validity of the walking scale for spinal cord injury and other domains of function in a multicenter clinical trial. *Neurorehabil Neural Repair.* 2007;21(6):539-550.
28. Donnelly C, Eng JJ, Hall J, et al. Client-centred assessment and the identification of meaningful treatment goals for individuals with a spinal cord injury. *Spinal Cord.* 2004;42(5):302-307.
29. Schober P, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and analgesia.* 2018;126(5):1763-1768.
30. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil.* 1993;74(5):531-536.
31. Spooren AI, Janssen-Potten YJ, Post MW, Kerckhofs E, Nene A, Seelen HA. Measuring change in arm hand skilled performance in persons with a cervical spinal cord injury: responsiveness of the Van Lieshout Test. *Spinal Cord.* 2006;44(12):772-779.
32. Itzkovich M, Gelernter I, Biering-Sorensen F, et al. The Spinal Cord Independence Measure (SCIM) version III: reliability and validity in a multi-center international study. *Disabil Rehabil.* 2007;29(24):1926-1933.
33. Middleton JW, Harvey LA, Batty J, Cameron I, Quirk R, Winstanley J. Five additional mobility and locomotor items to improve responsiveness of the FIM in wheelchair-dependent individuals with spinal cord injury. *Spinal Cord.* 2006;44(8):495-504.
34. Ottenbacher KJ, Hsu Y, Granger CV, Fiedler RC. The reliability of the functional independence measure: a quantitative review. *Arch Phys Med Rehabil.* 1996;77(12):1226-1232.
35. Segal ME, Ditunno JF, Staas WE. Interinstitutional agreement of individual functional independence measure (FIM) items measured at two sites on one sample of SCI patients. *Paraplegia.* 1993;31(10):622-631.

36. National Spinal Cord Injury Statistical Center, Facts and Figures At a Glance. Birmingham, AL: University of Alabama at Birmingham, 2019. https://www.nscisc.uab.edu/Public/Facts%20and%20Figures%202019%20-%20Final.pdf. Accessed 13 March, 2019.
37. Grey N, Kennedy P. The Functional Independence Measure: a comparative study of clinician and self ratings. *Paraplegia.* 1993;31(7):457-461.
38. Karamehmetoglu SS, Karacan I, Elbasi N, Demirel G, Koyuncu H, Dosoglu M. The functional independence measure in spinal cord injured patients: comparison of questioning with observational rating. *Spinal Cord.* 1997;35(1):22-25.
39. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ.* 1997;314(7080):572.
40. Kucukdeveci AA, Yavuzer G, Elhan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. *Clinical rehabilitation.* 2001;15(3):311-319.
41. Gage B, Deutsch A, Smith L, et al. The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on CARE Item Set and Current Assessment Comparisons. Research Triangle Institute. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Post-Acute-Care-Quality-Initiatives/Downloads/The-Development-and-Testing-of-the-Continuity-Assessment-Record-and-Evaluation-CARE-Item-Set-Final-Report-on-the-Development-of-the-CARE-Item-Set-and-Current-Assessment-Comparisons-Volume-3-of-3.pdf. Published 2012. Accessed January 5, 2019.
42. Catz A, Itzkovich M, Agranov E, Ring H, Tamir A. SCIM--spinal cord independence measure: a new disability scale for patients with spinal cord lesions. *Spinal Cord.* 1997;35(12):850-856.
43. Catz A, Itzkovich M, Steinberg F, et al. The Catz-Itzkovich SCIM: a revised version of the Spinal Cord Independence Measure. *Disabil Rehabil.* 2001;23(6):263-268.
44. Itzkovich M, Shefler H, Front L, et al. SCIM III (Spinal Cord Independence Measure version III): reliability of assessment by interview and comparison with assessment by observation. *Spinal Cord.* 2018;56(1):46-51.
45. Mulcahey MJ, Calhoun CL, Sinko R, Kelly EH, Vogel LC. The spinal cord independence measure (SCIM)-III self report for youth. *Spinal Cord.* 2016;54(3):204-212.
46. Catz A, Itzkovich M, Tesio L, et al. A multicenter international study on the Spinal Cord Independence Measure, version III: Rasch psychometric validation. *Spinal Cord.* 2007;45(4):275-291.
47. Rudhe C, van Hedel HJ. Upper extremity function in persons with tetraplegia: relationships between strength, capacity, and the spinal cord independence measure. *Neurorehabil Neural Repair.* 2009;23(5):413-421.
48. van Hedel HJ. Gait speed in relation to categories of functional ambulation after spinal cord injury. *Neurorehabil Neural Repair.* 2009;23(4):343-350.
49. Wirz M, Muller R, Bastiaenen C. Falls in persons with spinal cord injury: validity and reliability of the Berg Balance Scale. *Neurorehabil Neural Repair.* 2010;24(1):70-77.
50. Marino RJ, Kern SB, Leiby B, Schmidt-Read M, Mulcahey MJ. Reliability and validity of the capabilities of upper extremity test (CUE-T) in subjects with chronic spinal cord injury. *J Spinal Cord Med.* 2015;38(4):498-504.
51. Ackerman P, Morrison SA, McDowell S, Vazquez L. Using the Spinal Cord Independence Measure III to measure functional recovery in a post-acute spinal cord injury program. *Spinal Cord.* 2010;48(5):380-387.
52. Aidinoff E, Front L, Itzkovich M, et al. Expected spinal cord independence measure, third version, scores for various neurological levels after complete spinal cord lesions. *Spinal Cord.* 2011;49(8):893-896.

53. Scivoletto G, Tamburella F, Laurenza L, Molinari M. The spinal cord independence measure: how much change is clinically significant for spinal cord injury subjects. *Disabil Rehabil.* 2013;35(21):1808-1813.
54. Catz A, Itzkovich M, Agranov E, Ring H, Tamir A. The spinal cord independence measure (SCIM): sensitivity to functional changes in subgroups of spinal cord lesion patients. *Spinal Cord.* 2001;39(2):97-100.
55. Anderson K, Aito S, Atkins M, et al. Functional recovery measures for spinal cord injury: an evidence-based review for clinical practice and research. *J Spinal Cord Med.* 2008;31(2):133-144.
56. Fekete C, Eriks-Hoogland I, Baumberger M, et al. Development and validation of a self-report version of the Spinal Cord Independence Measure (SCIM III). *Spinal Cord.* 2013;51(1):40-47.
57. Spinal Cord Outcomes Partnership Endeavor, Current SCI Clinical Trials of Drug, Cell, and Surgical Interventions to Improve Neurological and Related Functional Outcomes. . http://www.scope-sci.org/trials.php. Updated Updated April 2015. Accessed.
58. Charlifue S, Tate D, Biering-Sorensen F, et al. Harmonization of Databases: A Step for Advancing the Knowledge About Spinal Cord Injury. *Arch Phys Med Rehabil.* 2016;97(10):1805-1818.
59. Kopp MA, Watzlawick R, Martus P, et al. Long-term functional outcome in patients with acquired infections after acute spinal cord injury. *Neurology.* 2017;88(9):892-900.
60. Cohen JT, Marino RJ, Sacco P, Terrin N. Association between the functional independence measure following spinal cord injury and long-term outcomes. *Spinal Cord.* 2012;50(10):728-733.
61. Putzke JD, Barrett JJ, Richards JS, DeVivo MJ. Age and spinal cord injury: an emphasis on outcomes among the elderly. *J Spinal Cord Med.* 2003;26(1):37-44.
62. Marino RJ, Burns S, Graves DE, Leiby BE, Kirshblum S, Lammertse DP. Upper- and lower-extremity motor recovery after traumatic cervical spinal cord injury: an update from the national spinal cord injury database. *Arch Phys Med Rehabil.* 2011;92(3):369-375.
63. Kirshblum S, Millis S, McKinley W, Tulsky D. Late neurologic recovery after traumatic spinal cord injury. *Arch Phys Med Rehabil.* 2004;85(11):1811-1817.
64. Lee BA, Leiby BE, Marino RJ. Neurological and functional recovery after thoracic spinal cord injury. *J Spinal Cord Med.* 2016;39(1):67-76.
65. Hupp M, Pavese C, Bachmann LM, Koller R, Schubert M. Electrophysiological Multimodal Assessments Improve Outcome Prediction in Traumatic Cervical Spinal Cord Injury. *J Neurotrauma.* 2018;35(24):2916-2923.
66. Reed R, Mehra M, Kirshblum S, et al. Spinal cord ability ruler: an interval scale to measure volitional performance after spinal cord injury. *Spinal Cord.* 2017.
67. van Silfhout L, Hosman AJF, Bartels R, et al. Ten Meters Walking Speed in Spinal Cord-Injured Patients: Does Speed Predict Who Walks and Who Rolls? *Neurorehabil Neural Repair.* 2017;31(9):842-850.
68. Zariffa J, Kramer JL, Fawcett JW, et al. Characterization of neurological recovery following traumatic sensorimotor complete thoracic spinal cord injury. *Spinal Cord.* 2011;49(3):463-471.
69. Steeves JD, Kramer JK, Fawcett JW, et al. Extent of spontaneous motor recovery after traumatic cervical sensorimotor complete spinal cord injury. *Spinal Cord.* 2011;49(2):257-265.
70. Kramer JL, Lammertse DP, Schubert M, Curt A, Steeves JD. Relationship between motor recovery and independence after sensorimotor-complete cervical spinal cord injury. *Neurorehabil Neural Repair.* 2012;26(9):1064-1071.

71. Steeves JD, Lammertse DP, Kramer JL, et al. Outcome Measures for Acute/Subacute Cervical Sensorimotor Complete (AIS-A) Spinal Cord Injury During a Phase 2 Clinical Trial. *Topics in spinal cord injury rehabilitation.* 2012;18(1):1-14.
72. Dorans NJ. Linking scores from multiple health outcome instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2007;16 Suppl 1:85-94.
73. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychological assessment.* 2014;26(2):513-527.
74. Tulsky DS, Kisala PA, Boulton AJ, et al. Determining a transitional scoring link between PROMIS(R) pediatric and adult physical health measures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2019;28(5):1217-1229.
75. Noonan VK, Cook KF, Bamer AM, Choi SW, Kim J, Amtmann D. Measuring fatigue in persons with multiple sclerosis: creating a crosswalk between the Modified Fatigue Impact Scale and the PROMIS Fatigue Short Form. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2012;21(7):1123-1133.
76. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology.* 2013;4:863.
77. Monsell SE, Dodge HH, Zhou XH, et al. Results From the NACC Uniform Data Set Neuropsychological Battery Crosswalk Study. *Alzheimer disease and associated disorders.* 2016;30(2):134-139.
78. Armstrong MJ, Duff-Canning S, Psych C, Kowgier M, Marras C. Independent application of montreal cognitive assessment/mini-mental state examination conversion. *Movement disorders : official journal of the Movement Disorder Society.* 2015;30(12):1710-1711.
79. Lawton M, Kasten M, May MT, et al. Validation of conversion between mini-mental state examination and montreal cognitive assessment. *Movement disorders : official journal of the Movement Disorder Society.* 2016;31(4):593-596.
80. Lawton M, Hu MT, Baig F, et al. Equating scores of the University of Pennsylvania Smell Identification Test and Sniffin' Sticks test in patients with Parkinson's disease. *Parkinsonism & related disorders.* 2016;33:96-101.
81. Tulsky DS, Kisala PA, Boulton AJ, et al. Determining a transitional scoring link between PROMIS(R) pediatric and adult physical health measures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2018.
82. Wang YC, Byers KL, Velozo CA. Validation of FIM-MDS crosswalk conversion algorithm. *Journal of rehabilitation research and development.* 2008;45(7):1065-1076.
83. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England).* 1986;1(8476):307-310.
84. Emons WH, Habibovic M, Pedersen SS. Prevalence of anxiety in patients with an implantable cardioverter defibrillator: measurement equivalence of the HADS-A and the STAI-S. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2019;28(11):3107-3116.
85. Victorson D, Schalet BD, Kundu S, et al. Establishing a common metric for self-reported anxiety in patients with prostate cancer: Linking the Memorial Anxiety Scale for Prostate Cancer with PROMIS Anxiety. *Cancer.* 2019;125(18):3249-3258.
86. Boulton AJ, Tyner CE, Choi SW, et al. Linking the GAD-7 and PHQ-9 to the TBI-QOL Anxiety and Depression Item Banks. *J Head Trauma Rehabil.* 2019;34(5):353-363.
87. Byers KL. *Testing the accuracy of linking healthcare data across the continuum of care*: Rehabilitation Science, University of Florida; 2004.

88. Ketchum JM, Whiteneck G, Eagye MA. Guidelines for Assessing Crosswalks. In:2016.
89. Velozo CA, Byers KL, Wang YC, Joseph BR. Translating measures across the continuum of care: using Rasch analysis to create a crosswalk between the Functional Independence Measure and the Minimum Data Set. *Journal of rehabilitation research and development.* 2007;44(3):467-478.
90. Williams BC, Li Y, Fries BE, Warren RL. Predicting patient scores between the functional independence measure and the minimum data set: development and performance of a FIM-MDS "crosswalk". *Arch Phys Med Rehabil.* 1997;78(1):48-54.
91. Kolen M BR. *Test Equating, Scaling and Linking: Methods and Practices.* Third ed. New York: Springer; 2014.
92. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel R. Clinical implications of Brief Psychiatric Rating Scale scores. *The British journal of psychiatry : the journal of mental science.* 2005;187:366-371.
93. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophrenia research.* 2005;79(2-3):231-238.
94. Leucht S, Kane JM, Etschel E, Kissling W, Hamann J, Engel RR. Linking the PANSS, BPRS, and CGI: clinical implications. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology.* 2006;31(10):2318-2325.
95. Leucht S, Engel RR, Davis JM, et al. Equipercentile linking of the Brief Psychiatric Rating Scale and the Clinical Global Impression Scale in a catchment area. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology.* 2012;22(7):501-505.
96. Levine SZ, Rabinowitz J, Engel R, Etschel E, Leucht S. Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophrenia research.* 2008;98(1-3):318-322.
97. Levine SZ, Leucht S. Identifying clinically meaningful symptom response cut-off values on the SANS in predominant negative symptoms. *Schizophrenia research.* 2013;145(1-3):125-127.
98. Schennach-Wolff R, Obermeier M, Seemuller F, et al. Does clinical judgment of baseline severity and changes in psychopathology depend on the patient population? Results of a CGI and PANSS linking analysis in a naturalistic study. *Journal of clinical psychopharmacology.* 2010;30(6):726-731.
99. Samara MT, Engel RR, Millier A, Kandenwein J, Toumi M, Leucht S. Equipercentile linking of scales measuring functioning and symptoms: examining the GAF, SOFAS, CGI-S, and PANSS. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology.* 2014;24(11):1767-1772.
100. Fong TG, Fearing MA, Jones RN, et al. Telephone interview for cognitive status: Creating a crosswalk with the Mini-Mental State Examination. *Alzheimer's & dementia : the journal of the Alzheimer's Association.* 2009;5(6):492-497.
101. Fong TG, Jones RN, Rudolph JL, et al. Development and validation of a brief cognitive assessment tool: the sweet 16. *Archives of internal medicine.* 2011;171(5):432-437.
102. Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimer's & dementia : the journal of the Alzheimer's Association.* 2013;9(5):529-537.
103. Roalf DR, Moore TM, Mechanic-Hamilton D, et al. Bridging cognitive screening tests in neurologic disorders: A crosswalk between the short Montreal Cognitive Assessment and Mini-Mental State Examination. *Alzheimer's & dementia : the journal of the Alzheimer's Association.* 2017;13(8):947-952.

104. van Steenoven I, Aarsland D, Hurtig H, et al. Conversion between mini-mental state examination, montreal cognitive assessment, and dementia rating scale-2 scores in Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society.* 2014;29(14):1809-1815.
105. Saczynski JS, Inouye SK, Guess J, et al. The Montreal Cognitive Assessment: Creating a Crosswalk with the Mini-Mental State Examination. *Journal of the American Geriatrics Society.* 2015;63(11):2370-2374.
106. Ghomrawi HM, Lee YY, Herrero C, et al. A Crosswalk Between UCLA and Lower Extremity Activity Scales. *Clinical orthopaedics and related research.* 2017;475(2):542-548.
107. ten Klooster PM, Oude Voshaar MA, Gandek B, et al. Development and evaluation of a crosswalk between the SF-36 physical functioning scale and Health Assessment Questionnaire disability index in rheumatoid arthritis. *Health and quality of life outcomes.* 2013;11:199.
108. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association.* 1999;53(1):83-90.
109. Linacre J. Help for winsteps rasch measurement software : Dimensionality: contrasts and variances. https://www.rasch.org/rmt/rmt201c.htm. Accessed December 2, 2018.
110. Wright B, Linacre J. Reasonable Mean Fit Square Values. *Rasch Measurement Transaction.* 1994;8(3):370.
111. Malec JF, Whiteneck GG, Bogner JA. Another Look at the PART-O Using the Traumatic Brain Injury Model Systems National Database: Scoring to Optimize Psychometrics. *Arch Phys Med Rehabil.* 2016;97(2):211-217.
112. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical care.* 2007;45(5 Suppl 1):S22-31.
113. Principal components and exploratory factor analysis with SPSS. UCLA Statistical Consulting Group. https://stats.idre.ucla.edu/spss/seminars/efa-spss/. Accessed December 2, 2018.
114. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. *J Rehabil Med.* 2011;43(10):884-891.
115. Linacre JM, Heinemann AW, Wright BD, Granger CV, Hamilton BB. The structure and stability of the Functional Independence Measure. *Arch Phys Med Rehabil.* 1994;75(2):127-132.
116. Dijkers MP, Yavuzer G. Short versions of the telephone motor Functional Independence Measure for use with persons with spinal cord injury. *Arch Phys Med Rehabil.* 1999;80(11):1477-1484.
117. Dijkers MP. A computer adaptive testing simulation applied to the FIM instrument motor component. *Arch Phys Med Rehabil.* 2003;84(3):384-393.
118. Lawton G, Lundgren-Nilsson A, Biering-Sorensen F, et al. Cross-cultural validity of FIM in spinal cord injury. *Spinal Cord.* 2006;44(12):746-752.
119. Kozlowski AJ, Heinemann AW. Using individual growth curve models to predict recovery and activities of daily living after spinal cord injury: an SCIRehab project study. *Arch Phys Med Rehabil.* 2013;94(4 Suppl):S154-164.e151-154.
120. Whiteneck G, Gassaway J, Dijkers M, et al. The SCIRehab project: treatment time spent in SCI rehabilitation. Inpatient treatment time across disciplines in spinal cord injury rehabilitation. *J Spinal Cord Med.* 2011;34(2):133-148.
121. Baglin J. Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation.* 2014;19(5):2.

122. Li CY, Romero S, Bonilha HS, et al. Linking Existing Instruments to Develop an Activity of Daily Living Item Bank. *Evaluation & the health professions.* 2016.
123. Hong I, Lim Y, Han H, Hay CC, Woo HS. Application of the Korean Version of the Modified Barthel Index: Development of a keyform for use in Clinical Practice. *Hong Kong journal of occupational therapy : HKJOT.* 2017;29(1):39-46.
124. American Spinal Injury Association: International Standards for Neurological Classification of Spinal Cord Injury, revised 2019; Richmond, VA. In.
125. Ketchum JM, Cuthbert JP, Deutsch A, et al. Representativeness of the Spinal Cord Injury Model Systems National Database. *Spinal Cord.* 2018;56(2):126-132.

APPENDIX

A. Spinal Cord Independence Measure Version III

שירותי בריאות כללית

LOEWENSTEIN HOSPITAL REHABILITATION CENTER
Affiliated with the Sackler Faculty of Medicine, Tel-Aviv University
Department IV, Medical Director: Dr. Amiram Catz **Tel:** 972-9-7709090 **Fax:** 972-9-7709986 e-mail: amiramc@clalit.org.il
Patient Name: _____ **ID:**_____ **Examiner Name:** _____
(Enter the score for each function in the adjacent square, below the date. The form may be used for up to 6 examinations.)

SCIM-SPINAL CORD INDEPENDENCE MEASURE

Version III, Sept 14, 2002

Self-Care DATE (Exam 1 2 3 4 5 6)

1. Feeding (cutting, opening containers, pouring, bringing food to mouth, holding cup with fluid)
0. Needs parenteral, gastrostomy, or fully assisted oral feeding
1. Needs partial assistance for eating and/or drinking, or for wearing adaptive devices
2. Eats independently; needs adaptive devices or assistance only for cutting food and/or pouring and/or opening containers
3. Eats and drinks independently; does not require assistance or adaptive devices

2. Bathing (soaping, washing, drying body and head, manipulating water tap). **A-upper body; B-lower body**
A. 0. Requires total assistance
1. Requires partial assistance
2. Washes independently with adaptive devices or in a specific setting (e.g., bars, chair)
3. Washes independently; does not require adaptive devices or specific setting (not customary for healthy people) (adss)
B. 0. Requires total assistance
1. Requires partial assistance
2. Washes independently with adaptive devices or in a specific setting (adss)
3. Washes independently; does not require adaptive devices (adss) or specific setting

3. Dressing (clothes, shoes, permanent orthoses: dressing, wearing, undressing). **A-upper body; B-lower body**
A. 0. Requires total assistance
1. Requires partial assistance with clothes without buttons, zippers or laces (cwobzl)
2. Independent with cwobzl; requires adaptive devices and/or specific settings (adss)
3. Independent with cwobzl; does not require adss; needs assistance or adss only for bzl
4. Dresses (any cloth) independently; does not require adaptive devices or specific setting
B. 0. Requires total assistance
1. Requires partial assistancewith clothes without buttons, zipps or laces (cwobzl)
2. Independent with cwobzl; requires adaptive devices and/or specific settings (adss)
3. Independent with cwobzl without adss; needs assistance or adss only for bzl
4. Dresses (any cloth) independently; does not require adaptive devices or specific setting

4. Grooming (washing hands and face, brushing teeth, combing hair, shaving, applying makeup)
0. Requires total assistance
1. Requires partial assistance
2. Grooms independently with adaptive devices
3. Grooms independently without adaptive devices

SUBTOTAL (0-20)

Respiration and Sphincter Management

5. Respiration
0. Requires tracheal tube (TT) and permanent or intermittent assisted ventilation (IAV)
2. Breathes independently with TT; requires oxygen, much assistance in coughing or TT management
4. Breathes independently with TT; requires little assistance in coughing or TT management
6. Breathes independently without TT; requires oxygen, much assistance in coughing, a mask (e.g., peep) or IAV (bipap)
8. Breathes independently without TT; requires little assistance or stimulation for coughing
10. Breathes independently without assistance or device

6. Sphincter Management - Bladder
0. Indwelling catheter
3. Residual urine volume (RUV) > 100cc; no regular catheterization or assisted intermittent catheterization
6. RUV < 100cc or intermittent self-catheterization; needs assistance for applying drainage instrument
9. Intermittent self-catheterization; uses external drainage instrument; does not need assistance for applying
11. Intermittent self-catheterization; continent between catheterizations; does not use external drainage instrument
13. RUV <100cc; needs only external urine drainage; no assistance is required for drainage
15. RUV <100cc; continent; does not use external drainage instrument

7. Sphincter Management - Bowel
0. Irregular timing or very low frequency (less than once in 3 days) of bowel movements
5. Regular timing, but requires assistance (e.g., for applying suppository); rare accidents (less than twice a month)
8. Regular bowel movements, without assistance; rare accidents (less than twice a month)
10. Regular bowel movements, without assistance; no accidents

8. Use of Toilet (perineal hygiene, adjustment of clothes before/after, use of napkins or diapers).
0. Requires total assistance
1. Requires partial assistance; does not clean self
2. Requires partial assistance; cleans self independently
4. Uses toilet independently in all tasks but needs adaptive devices or special setting (e.g., bars)
5. Uses toilet independently; does not require adaptive devices or special setting)

SUBTOTAL (0-40)

Mobility (room and toilet) DATE \ \ \ \ \ \ \

9. Mobility in Bed and Action to Prevent Pressure Sores

0. Needs assistance in all activities: turning upper body in bed, turning lower body in bed, sitting up in bed, doing push-ups in wheelchair, with or without adaptive devices, but not with electric aids
2. Performs one of the activities without assistance
4. Performs two or three of the activities without assistance
6. Performs all the bed mobility and pressure release activities independently

10. Transfers: bed-wheelchair (locking wheelchair, lifting footrests, removing and adjusting arm rests, transferring, lifting feet).

0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., sliding board)
2. Independent (or does not require wheelchair)

11. Transfers: wheelchair-toilet-tub (if uses toilet wheelchair: transfers to and from; if uses regular wheelchair: locking wheelchair, lifting footrests, removing and adjusting armrests, transferring, lifting feet)

0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., grab-bars)
2. Independent (or does not require wheelchair)

Mobility (indoors and outdoors, on even surface)

12. Mobility Indoors

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal walking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

13. Mobility for Moderate Distances (10-100 meters)

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal walking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

14. Mobility Outdoors (more than 100 meters)

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal waking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

15. Stair Management

0. Unable to ascend or descend stairs
1. Ascends and descends at least 3 steps with support or supervision of another person
2. Ascends and descends at least 3 steps with support of handrail and/or crutch or cane
3. Ascends and descends at least 3 steps without any support or supervision

16. Transfers: wheelchair-car (approaching car, locking wheelchair, removing arm- and footrests, transferring to and from car, bringing wheelchair into and out of car)

0. Requires total assistance
1. Needs partial assistance and/or supervision and/or adaptive devices
2. Transfers independent; does not require adaptive devices (or does not require wheelchair)

17. Transfers: ground-wheelchair

0. Requires assistance
1. Transfers independent with or without adaptive devices (or does not require wheelchair)

SUBTOTAL (0-40)

TOTAL SCIM SCORE (0-100)

APPENDIX

B. Formulas

Reduction in uncertainty	$RiU = 1 - CoA = 1 - \sqrt{1 - r^2}$	$RiU = Reduction\ in\ uncertainty$ $CoA =$ coefficient of alienation $r =$ correlation coefficient
Rasch analysis	$\log\left[\frac{P_{nijk}}{p_{nijk-1}}\right] = B_n - D_i - C_j - F_k$	$P_{nijk} =$ probability of person passing a test item $p_{nijk-1} =$ probability of a person not passing a test item $B_n =$ person ability $D_i =$ item difficulty $C_j =$ rater (judge)severity $F_k =$ difficulty of the rating step
Pearson correlation coefficient	$r = \frac{N\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$	$N =$ number of paired scores $\Sigma xy =$ sum of products of paired scores $\Sigma x =$ sum of x scores $\Sigma y =$ sum of y scores Σx^2=sum of squared x scores Σy^2=sum of squared y scores
Cohen's effect size and Standardized mean difference (used for sub-group invariance)	Cohen's $D = \frac{(M_2 - M_1)}{SD_{pooled}}$ Where $SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$	$M_2 =$ mean in group 2 $M_1 =$ mean in group 1 $SD_1 =$ standard deviation group 1 $SD_2 =$ standard deviation group 2

APPENDIX

C. Coding

1.0 Winsteps Rasch coding for co-calibration

Run 1: Analyze FIM items separately.
Run 2: Analyze SCIM items separately.
Run 3: Co-calibrate FIM and SCIM using group rating scale model and analyze jointly
Run 4: Anchor FIM to item and rating scale measures anchor file
Run 5: Anchor SCIM to item and rating measures anchor files

Where no run is indicated, line of code used for each run. Where indicated in quotes, line of code used

for specific runs, as indicated by number.

```
&INST

Title = "FIM-SCIM 27 Items for Rasch.xlsx" ; title of file with FIM SCIM data
 ; Sheet1
 ;Excel Cases processed = 662 ; number of data points
; Excel Variables processed = 28 ; number of variables in this case subject ID + 27 FIM/SCIM items
 ITEM1 = 1 ; Starting column of item responses ; first column with data – in this case subject ID
 NI = 27 ; Number of items; number of columns with data, 11 FIM + 16 SCIM
NAME1 = 29 ; Starting column for person label in data record
NAMLEN = 4 ; Length of person label; # maximum digits/letters in subject ID
 XWIDE = 1 ; Matches the widest data value observed;
; GROUPS = 0 ; Partial Credit model: in case items have different rating scales; not used in this analysis
“Run 2,5” IDELETE=1-11; tells Winsteps to ignore first 11 items which in this case are FIM items
“Run 1,4” IDELETE=12-27; tells Winsteps to ignore items 12-27 which in this case are SCIM items


CODES = "012345678"    ; valid data codes
IVALUEA = "*1234567*"   ; all FIM items
IVALUEB = "0123*****"   ; SFeed, , SBathUB, SBathLB, SGroom, SStair
IVALUEC = "01234****"   ; SDressUB, SDressLB
IVALUED = "012*45***"   ; SToilet
IVALUEE = "012******"   ; SXferBed, SXferToil, SXferCar
IVALUEF = "012345678"   ; SMobInd, SModDist, SMobOut
IVALUEG = "0*2*4*6**"   ; SMobBed
IVALUEH = "01*******"   ; SXferGRND

 STKEEP = YES
```

```
            ;0        1         2                  0=1; 1=10, 2=20
        ;12345678901234567890123456789
  IREFER = AAAAAAAAAAABBBCCBDEEFFFBGEH      ;items to rescore
  GROUPS =AAAAAAAAAAABBBCCBDEEFFFBGEH       ;item groupings

  TOTALSCORE = Yes ; Include extreme responses in reported scores
; Person Label variables: columns in label: columns in line
@SubID = 1E3 ; $C29W3

"Run 4,5" IAFILE= C:\Winsteps\Winsteps_Linda_ partial rating\IFILE_FIM-SCIM.txt ; tells Winsteps to link
FIM or SCIM to FIM/SCIM item anchor file
"Run 4,5" SAFILE= C:\Winsteps\Winsteps_Linda_ partial rating\SFILE FIM-SCIM.txt; tells Winsteps to link
FIM or SCIM to FIM/SCIM item structure anchor file

&END ; Item labels follow: columns in label
FEat ; Item 1 : 1-1
FGroom ; Item 2 : 2-2
FBath ; Item 3 : 3-3
FDressUB ; Item 4 : 4-4
FDressLB ; Item 5 : 5-5
FToilet ; Item 6 : 6-6
FXfrBed ; Item 7 : 7-7
FXfrToil ; Item 8 : 8-8
FXfrTub ; Item 9 : 9-9
FWalkwc ; Item 10 : 10-10
FStairs ; Item 11 : 11-11
SFeed ; Item 12 : 12-12
SBathUB ; Item 13 : 13-13
SBathLB ; Item 14 : 14-14
SDressUB ; Item 15 : 15-15
SDressLB ; Item 16 : 16-16
SGroom ; Item 17 : 17-17
SToilet ; Item 18 : 18-18
SXfrBed ; Item 19 : 19-19
SXfrToil ; Item 20 : 20-20
SMobInd ; Item 21 : 21-21
SModDist ; Item 22 : 22-22
SOut ; Item 23 : 23-23
SStair ; Item 24 : 24-24
SMobBed ; Item 25 : 25-25
SXfrCar ; Item 26 : 26-26
SXfrGrnd ; Item 27 : 27-27
```

1.1 SPSS expert panel, equipercentile and Rasch recoding

Expert panel FIM recoding:

```
RECODE FIMEat (7=3) (0 thru 1=0) (2 thru 4=1) (5 thru 6=2) INTO EFSFIMEat.
VARIABLE LABELS  EFSFIMEat 'Recode FIM Eat to EFS Eat'.
EXECUTE.

RECODE FIMBath (6=2) (7=3) (0 thru 1=0) (2 thru 5=1) INTO EFSFIMBath.
VARIABLE LABELS  EFSFIMBath 'EFS FIM Bath'.
EXECUTE.

RECODE FIMDressUB (7=3) (0 thru 1=0) (2 thru 4=1) (5 thru 6=2) INTO EFSFIMDressUB.
VARIABLE LABELS  EFSFIMDressUB 'EFS FIM Dress UB'.
EXECUTE.

RECODE FIMDressLB (7=3) (0 thru 1=0) (2 thru 4=1) (5 thru 6=2) INTO EFSFIMDressLB.
VARIABLE LABELS  EFSFIMDressLB 'EFS FIM Dress LB'.
EXECUTE.

RECODE FIMGroom (6=2) (7=3) (0 thru 1=0) (2 thru 5=1) INTO EFSFIMGroom.
VARIABLE LABELS  EFSFIMGroom 'EFS FIM Groom'.
EXECUTE.

RECODE FIMToilet (6=2) (7=3) (0 thru 1=0) (2 thru 5=1) INTO EFSFIMUsingToilet.
VARIABLE LABELS  EFSFIMUsingToilet 'EFS FIM Using Toilet'.
EXECUTE.


RECODE FIMXferBed (7=2) (0 thru 1=0) (2 thru 6=1) INTO EFSFIMXferBed.
VARIABLE LABELS  EFSFIMXferBed 'EFS FIM Xfer Bed to w/c'.
EXECUTE.

COMPUTE EFSFIMXferToilet=1.
EXECUTE.

IF  (FIMXferToilet < 2) and (FIMXferTub < 2) EFSFIMXferToilet=0.
EXECUTE.

IF  ((FIMXferToilet = 7) and (FIMXferTub = 7) ) EFSFIMXferToilet=2.
EXECUTE.

DO IF  (FIMfmodWalkwc=2).
RECODE FIMWalkwc (0 thru 1=0) (2 thru 7=1) INTO EFSFIMWheelWalk.
END IF.
VARIABLE LABELS  EFSFIMWheelWalk 'FIM locomotion to EFS '.
```

```
EXECUTE.

*FIMmodWalkwc of 2=wc

DO IF  ((FIMfmodWalkwc = 1) or (FIMfmodWalkwc = 3)).
RECODE FIMWalkwc (6=3) (7=4) (0 thru 1=0) (2 thru 5=2) INTO EFSFIMWheelWalk.
END IF.
VARIABLE LABELS  EFSFIMWheelWalk 'FIM locomotion to EFS '.
EXECUTE.

* FIM modWalkWc of 1=walk,2=w/c, 3=both

DATASET ACTIVATE DataSet1.
RECODE FIMStairs (6=2) (7=3) (0 thru 1=0) (2 thru 5=1) INTO EFSFIMStairs.
VARIABLE LABELS  EFSFIMStairs 'EFS FIM Stairs'.
EXECUTE.
```

Expert panel SCIM recoding:

```
RECODE SCIMFeed (0=0) (1=1) (2=2) (3=3) INTO EFSSCIMEat.
VARIABLE LABELS  EFSSCIMEat 'EFS SCIM Eat'.
EXECUTE.

COMPUTE EFSSCIMBathing=1.
EXECUTE.

IF  (SCIMBathUB = 2 or SCIMBathUB = 3) and (SCIMBathLB = 2 or SCIMBathLB = 3) EFSSCIMBathing=2.
EXECUTE.

IF  (SCIMBathUB = 0 and SCIMBathLB = 0) EFSSCIMBathing=0.
EXECUTE.

IF  (SCIMBathUB = 3 and SCIMBathLB =3) EFSSCIMBathing=3.
EXECUTE.

RECODE SCIMDressUB (0=0) (1=1) (4=3) (2 thru 3=2) INTO EFSSCIMDressUB.
VARIABLE LABELS  EFSSCIMDressUB 'EFS SCIM Dress UB'.
EXECUTE.

RECODE SCIMDressLB (0=0) (1=1) (4=3) (2 thru 3=2) INTO EFSSCIMDressLB.
VARIABLE LABELS  EFSSCIMDressLB 'EFS SCIM Dress LB'.
EXECUTE.


RECODE SCIMGroom (0=0) (1=1) (2=2) (3=3) INTO EFSSCIMGroom.
VARIABLE LABELS  EFSSCIMGroom 'EFS SCIM Groom'.
```

```
EXECUTE.

RECODE SCIMToilet (0=0) (4=2) (5=3) (1 thru 2=1) INTO EFSSCIMUsingToilet.
VARIABLE LABELS  EFSSCIMUsingToilet 'EFS SCIM Using Toilet'.
EXECUTE.

RECODE SCIMXferBed (0=0) (1=1) (2=2) INTO EFSSCIMXferBed.
VARIABLE LABELS  EFSSCIMXferBed 'EFS FIM Xfer Bed'.
EXECUTE.

RECODE SCIMXferToilet (0=0) (1=1) (2=2) INTO EFSSCIMXferToilet.
VARIABLE LABELS  EFSSCIMXferToilet 'EFS SCIM Xfer Toilet'.
EXECUTE.

RECODE SCIMMobDist (0=0) (3=2) (8=4) (1 thru 2=1) (4 thru 7=3) INTO EFSSCIMWheelWalk.
VARIABLE LABELS  EFSSCIMWheelWalk 'EFS SCIM Wheel/Walk'.
EXECUTE.

RECODE SCIMModDist (0=0) (6=3) (7=4) (1 thru 2=1) (2 thru 5=2) INTO EFSSCIMMobMod. VARIABLE
LABELS  EFSSCIMMobMod 'SCIM mobility moderate distance to EFS'. EXECUTE.

RECODE SCIMStair (0=0) (1=1) (2=2) (3=3) INTO EFSSCIMStairs.
VARIABLE LABELS  EFSSCIMStairs 'EFS SCIM Stairs'.
EXECUTE.
```

Equipercentile SCIM to FIM recoding

```
RECODE SCIMVolMotorSUM (0=11.10) (1=12.00) (2=13.00) (3=15.06) (4=16.33) (5=18.36) (6=19.67)
   (7=20.86) (8=22.14) (9=23.33) (10=24.06) (11=25.30) (12=26.69) (13=28.40) (14=29.67) (15=30.92)
   (16=32.00) (17=32.80) (18=34.47) (19=36.44) (20=37.72) (21=39.38) (22=41.30) (23=42.40) (24=43.25)
   (25=44.22) (26=45.14) (27=46.89) (28=48.00) (29=48.83) (30=50.20) (31=51.67) (32=52.71) (33=53.82)
   (34=54.33) (35=56.53) (36=58.33) (37=59.94) (38=62.07) (39=63.20) (40=64.55) (41=65.38) (42=66.00)
   (43=66.00) (44=67.00) (45=67.00) (46=67.80) (47=68.00) (48=68.00) (49=68.67) (50=69.00) (51=69.00)
   (52=69.50) (53=70.00) (54=70.00) (55=70.00) (56=70.50) (57=71.00) (58=71.67) (59=72.00) (60=73.00)
   (61=73.67) (62=74.00) (63=75.00) (64=75.56) (65=76.67) INTO EQFIM.
VARIABLE LABELS  EFIM 'Equipercentile SCIM to EQFIM'.
EXECUTE.
```

Equipercentile SCIM to FIM recoding

```
RECODE FIMVolMotorSUM (11=0.00) (12=0.83) (13=2.00) (14=2.86) (15=3.00) (16=3.67) (17=4.00)
   (18=5.00) (19=5.50) (20=6.17) (21=7.08) (22=8.00) (23=8.75) (24=9.74) (25=10.93) (26=11.45)
   (27=12.08) (28=13.00) (29=13.40) (30=14.25) (31=15.00) (32=15.94) (33=17.11) (34=18.00) (35=18.00)
   (36=18.90) (37=19.46) (38=20.00) (39=20.91) (40=21.00) (41=22.00) (42=22.67) (43=23.60) (44=24.78)
```

(45=25.75) (46=26.75) (47=27.00) (48=27.91) (49=29.17) (50=30.00) (51=30.57) (52=31.20) (53=32.29)
(54=33.40) (55=34.25) (56=35.00) (57=35.00) (58=36.00) (59=36.67) (60=37.00) (61=37.38) (62=38.00)
(63=38.67) (64=39.71) (65=40.45) (66=41.92) (67=44.63) (68=47.07) (69=50.27) (70=53.93) (71=57.13)
(72=58.43) (73=60.40) (74=61.20) (75=63.80) (76=64.29) (77=65.00) INTO EQSCIM.
VARIABLE LABELS ESCIM 'Equipercentile FIM to EQSCIM'.
EXECUTE.

Rasch SCIM to FIM recoding

RECODE SCIMVolMotorSUM (0=11) (1=11) (2=12) (3=13) (4=14) (5=15) (6=17) (7=18) (8=20) (9=22)
(10=23) (11=25) (12=26) (13=28) (14=29) (15=31) (16=32) (17=33) (18=35) (19=36) (20=38) (21=39.5)
(22=41) (23=42.5) (24=44) (25=46) (26=47) (27=49) (28=50) (29=52) (30=53) (31=54) (32=56) (33=57)
(34=58) (35=59) (36=60) (37=61) (38=62) (39=63) (40=63) (41=64) (42=65) (43=65) (44=66) (45=66)
(46=67) (47=67) (48=68) (49=68) (50=69) (51=69) (52=70) (53=70) (54=70) (55=71) (56=71) (57=72)
(58=72) (59=73) (60=73) (61=74) (62=74) (63=75) (64=76) (65=77) INTO RFIM.
VARIABLE LABELS RFIM 'Rasch raw SCIM to FIM'.
EXECUTE.

Rasch FIM to SCIM recoding

RECODE FIMVolMotorSUM (11=1) (12=2) (13=3) (14=4) (15=5) (16=6) (17=6) (18=7) (19=8) (20=8)
(21=9) (22=9) (23=10) (24=11) (25=11) (26=12) (27=12.5) (28=13) (29=14) (30=15) (31=15) (32=16)
(33=17) (34=17) (35=18) (36=19) (37=19) (38=20) (39=21) (40=21) (41=22) (42=23) (43=23) (44=24)
(45=24.5) (46=25) (47=26) (48=26) (49=27) (50=28) (51=28.5) (52=29) (53=30) (54=31) (55=31.5)
(56=32) (57=33) (58=34) (59=35) (60=36) (61=37) (62=38) (63=39) (64=41) (65=43) (66=44) (67=46)
(68=49) (69=51) (70=53) (71=55) (72=57) (73=60) (74=62) (75=63) (76=64) (77=65) INTO RSCIM.
VARIABLE LABELS RSCIM 'Raw FIM to RSCIM'.
EXECUTE.

APPENDIX

D. Details of Expert Panel FIM/SCIM III (EFS) crosswalk by item and score

SCIM - Feeding

0. Needs parenteral, gastrostomy or fully assisted oral feeding
1. Needs partial assistance for eating and/or drinking, or for wearing adaptive devices
2. Eats independently; needs adaptive devices or assistance only for cutting food and/or pouring and/or opening containers
3. Eats and drinks independently; does not require assistance or adaptive devices

EFS - Eating

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0, 1
1	Partial assistance SCIM=1 FIM=2-4
2	Independent with equipment or set-up SCIM=2 FIM=5,6
3	Independent w/o equipment or set-up SCIM=3 FIM=7

FIM - Eating

HELPER
0. Activity Does Not Occur— The patient does not eat and does not receive any parenteral/enteral nutrition during the entire assessment time frame. 1. Total Assistance- The patient performs less than 25% of eating tasks, or the patient relies on parenteral or gastrostomy feedings (either wholly or partially) and does not self-administer feedings.
2. Maximal assistance – The patient performs 25-49% of eating tasks.
3. Moderate Assistance – The patient performs 50-74% of eating tasks.
4. Minimal Contact Assistance- The patient performs 75% or more of eating tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by, cuing or coaxing) or setup (application of orthoses or assistive/adaptive devices), or another person is required to open containers, butter bread, cut meat, or pour liquids.
NO HELPER
6. Modified Independence – Performance of the activity involves safety considerations, or the patient requires adaptive or assistive device such as long straw, spork, or rocking knife; requires more than a reasonable time to eat; or requires modified food consistency or blenderized food. If the patient relies on other means of alimentation, such as parenteral or gastrostomy feedings, then (s)he self administers feedings.
7. Complete Independence – The patient eats from a dish while managing a variety of food consistencies, and drinks from a cup or glass with the meal presented in the customary manner on a table or tray. The subject opens containers, butters bread, cuts meat, pours liquids, and uses a spoon or fork to bring food to the mouth, where it is chewed and swallowed. The patient performs this activity safely.

SCIM – Bathing Upper Body and Lower Body items

0. Requires total assistance
1. Requires partial assistance
2. Washes independently with adaptive devices or in a specific setting (e.g. bars, chair)
3. Washes independently; does not require adaptive devices or a specific setting (not customary to healthy people)

EFS – Bathing (Upper and Lower Body)

#	Description and FIM/SCIM scores
0	Total assistance SCIM (both upper and lower body)=0 FIM=0,1
1	Partial assistance SCIM (upper and lower body)= all other combinations not specified for EFS=0, 2, and 3 FIM=2-5
2	Independent with equipment SCIM (upper and lower body)=2&2, 2&3, or 3&2 FIM=6
3	Independent w/o equipment SCIM (both upper and lower body)=3 FIM=7

FIM - Bathing

HELPER
0. Activity does not occur. Patient does not bathe self, and is not bathed by a helper. 1. Total Assistance – The patient performs less than 25% of bathing tasks.
2. Maximal Assistance - The patient performs 25-49% of bathing tasks.
3. Moderate Assistance – The patient performs 50-74% of eating tasks
4. Minimal Contact Assistance- The patient performs 75% or more of bathing tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by, cuing or coaxing) or setup (application of assistive/adaptive devices, setting out bathing equipment, and initial preparation such as preparing the water or washing materials).
NO HELPER
6. Modified Independence –The patient requires specialized equipment (including prosthesis or orthosis) to bathe, or takes more than a reasonable amount of time, or there are safety considerations.
7. Complete Independence – The patient safely bathes (washes, rinses and dries) the body.

SCIM – Dressing upper body

0. Requires total assistance
1. Requires partial assistance with clothes without buttons, zippers or laces
2. Independent with clothes without buttons, zippers or laces; requires adaptive devices and/or in a specific setting
3.. Independent with clothes without buttons, zippers or laces ; does not require adaptive devices and/or in a specific setting ; requires assistance or adaptive devices and/or in a specific setting only for buttons, zippers or laces
4. Dresses (any cloth) independently; does not require adaptive devices or specific setting

EFS – Dressing upper body

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0,1
1	Partial assistance SCIM=1 FIM=2,3, 4
2	Independent except fasteners or with equipment SCIM=2,3 FIM=5,6
3	Independent w/o equipment SCIM=4 FIM=7

FIM – Dressing upper body

HELPER
0. Activity Does Not Occur— The patient does not dress and the helper does not dress the patient in clothing that is appropriate to wear in public during the entire assessment time frame. The subject who wears only a hospital gown would be coded "0 – Activity Does Not Occur." Putting on and taking off scrubs may be appropriate for purposes of assessment. 1. Total Assistance – The patient performs less than 25% of dressing tasks.
2. Maximal Assistance - The patient performs 25-49% of dressing tasks.
3. Moderate Assistance – The patient performs 50-74% of dressing tasks.
4.Minimal Contact Assistance- The patient performs 75% or more of dressing tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by, cuing or coaxing) or setup (application of an upper or lower body or limb orthosis/prosthesis, application of an assistive/adaptive device, or setting out clothes or dressing equipment).
NO HELPER
6. Modified Independence – The patient requires special adaptive closure such as a Velcro fastener, or an assistive device (including a prosthesis or orthosis) to dress, or takes more than a reasonable amount of time.
7. Complete Independence — The patient dresses and undresses self. This includes obtaining clothes from their customary places (such as drawers and closets), and may include managing a bra, pullover garment, front-opening garment, zippers, buttons, or snaps, as well as the application and removal of a prosthesis or orthosis (which is not used as an assistive device for upper body dressing) when applicable. The patient performs this activity safely.

SCIM – Dressing lower body

0. Requires total assistance
1. Requires partial assistance with clothes without buttons, zippers or laces
2. Independent with clothes without buttons, zippers or laces; requires adaptive devices and/or in a specific setting
3. Independent with clothes without buttons, zippers or laces ; does not require adaptive devices and/or in a specific setting ; requires assistance or adaptive devices and/or in a specific setting only for buttons, zippers or laces
4. Dresses (any cloth) independently; does not require adaptive devices or specific setting

EFS - Dressing lower body

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0,1
1	Partial assistance SCIM=1 FIM=2,3,4
2	Independent except fasteners or with equipment SCIM=2,3 FIM=5,6
3	Independent w/o equipment SCIM=4 FIM=7

FIM – Dressing lower body

HELPER
0. Activity Does Not Occur— The patient does not dress and the helper does not dress the patient in clothing that is appropriate to wear in public during the entire assessment time frame. For example, the patient who wears only a hospital gown and/or underpants and/or footwear would be coded "0 – Activity Does Not Occur" for this item. Putting on and taking off scrubs may be appropriate for purposes of assessment. 1. Total Assistance – The patient performs less than 25% of dressing tasks.
2. Maximal Assistance - The patient performs 25-49% of dressing tasks.
3. Moderate Assistance – The patient performs 50-74% of dressing tasks.
4.Minimal Contact Assistance- The patient performs 75% or more of dressing tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by, cuing or coaxing) or setup (application of an upper or lower body or limb orthosis/prosthesis, application of an assistive/adaptive device, or setting out of clothes or dressing equipment).
NO HELPER
6. Modified Independence – The patient requires special adaptive closure such as a Velcro fastener, or an assistive device (including a prosthesis or orthosis) to dress, or takes more than a reasonable amount of time.
7. Complete Independence – The patient dresses and undresses self. This includes obtaining clothes from their customary places (such as drawers and closets), and may include managing underpants, slacks, skirt, belt, stockings, shoes, zippers, buttons, and snaps, as well as the application and removal of a prosthesis or orthosis (which is not used as an assistive device for lower body dressing) when applicable.

SCIM - Grooming

0. Requires total assistance
1. Requires partial assistance
2. Grooms independently with adaptive devices
3. Grooms independently without adaptive devices

EFS - Grooming

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0,1
1	Partial assistance SCIM=1 FIM=2-5
2	Independent with equipment SCIM=2 FIM=6
3	Independent w/o equipment SCIM=3 FIM=7

FIM - Grooming

HELPER
0. Activity Does Not Occur—The patient does not perform any grooming activities (oral care, hair grooming, washing the hands, washing the face, and either shaving the face or applying make-up), and is not groomed by a helper during the entire assessment time frame. 1. Total Assistance-The patient performs less than 25% or grooming tasks.
2. Maximal Assistance - The patient performs 25-49% of grooming tasks.
3. Moderate Assistance – The patient performs 50-74%, or more of grooming tasks.
4. Minimal Contact Assistance- The patient performs 75% or more of grooming tasks.
Supervision or Setup—The patient requires supervision (e.g., standing by, cuing, or coaxing) or setup (application of orthoses or adapted/assistive devices, setting out grooming equipment, or initial preparation such as applying toothpaste to toothbrush or opening make-up containers).
NO HELPER
6. Modified Independence- The patient requires specialized equipment (including prosthesis or orthosis) to perform grooming activities, or takes more than a reasonable amount of time, or there are safety considerations.
7. Complete Independence—The patient cleans teeth or dentures, combs or brushes hair, washes the hands, washes the face, and either shaves the face or applies make-up, including all preparations. The patient performs this activity safely.

SCIM – Use of Toilet

0. Requires total assistance
1. Requires partial assistance; does not clean self
2. Requires partial assistance; cleans self independently
4. Uses toilet independently in all tasks but needs adaptive devices or special setting (e.g., bars)
5. Uses toilet independently; does not require adaptive devices or special setting (e.g., bars)

EFS - Using Toilet

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0,1
1	Partial assistance SCIM=1,2 FIM=2-5
2	Independent with equipment SCIM=4 FIM=6
3	Independent w/o equipment SCIM=5 FIM=7

FIM – Toileting

HELPER
0. Activity does not occur. The patient does not perform *any* of the toileting tasks (perineal cleansing, clothing adjustment before and after toilet use, etc.) and helper does not perform any of these activities for the subject. 1. Total Assistance – The patient performs less than 25% of toileting tasks.
2. Maximal Assistance - The patient performs 25-49% of toileting tasks.
3. Moderate Assistance – The patient performs 50-74% of toileting tasks.
4.Minimal Contact Assistance- The patient performs 75% or more of toileting tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by, cuing or coaxing) or setup (application of adaptive devices or opening of packages)
NO HELPER
6. Modified Independence—The patient requires specialized equipment (including orthosis or prosthesis) during toileting, or takes more than a reasonable amount of time, or there are safety considerations.
7. Complete Independence—The patient safely cleanses self after voiding and bowel movements, and safely adjusts clothing before and after using toilet, bedpan, commode or urinal.

SCIM – Transfers: bed-wheelchair

0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., grab-bars)
2. Independent (or does not require wheelchair)

EFS – Transferring: bed-wheelchair

#	Description and FIM/SCIM scores
0	Total assistance SCIM=0 FIM=0,1
1	Partial assistance or with equipment SCIM=1 FIM=2-6
2	Independent w/o equipment SCIM=2 FIM=7

EFS – Transfers: bed, chair, wheelchair

HELPER
0. Activity Does Not Occur- The subject does not transfer to or from the bed or chair, and is not transferred to or from the bed or a chair by a helper or lifting device. 1. Total Assistance- The patient performs less than 25% of transferring task.
2. Maximal Assistance – The patient performs 25-49% of transferring tasks.
3. Moderate Assistance – The patient requires more help than touching or performs 50-74% of transferring tasks
4. Minimal Contact Assistance – The patient requires no more than touching and performs 75% or more of transferring tasks.
5. Supervision or Setup- The patient requires supervision (e.g., standing by or coaxing) or setup (positioning slide board, moving foot rests, etc.)
NO HELPER
6. Modified Independence: The patient requires an adaptive or assistive device such as sliding board, a lift, grab bars, or a special seat/braces/crutches: or the activity takes more than a reasonable amount of time; or there are safety considerations. In this case, A prosthesis or orthosis is considered an assistive device if used for the transfer.
7. Complete Independence – *If walking*, patient safely approaches, sits down on a regular chair, and gets up to a standing position from a regular chair. Patient also safely transfers from bed to chair. *If in a wheelchair*, patient approaches bed or chair, locks brakes, lifts foot rest if necessary, and performs either a standing pivot or sliding transfer (without a board) and returns. The patient performs this activity safely.

SCIM – Mobility in Bed and Action to Prevent Bed Sores

0. Needs assistance in all activities: turning upper body in bed, turning lower body in bed, sitting up in bed, doing push-ups in wheelchair, with or without adaptive devices, but not with electric aids
2. Performs one of the activities without assistance
4. Performs two or three of the activities without assistance
6. Performs all the bed mobility and pressure release activities independently

EFS

#	Description and FIM/SCIM scores
	Not able to map since no comparable items in FIM

FIM

SCIM – Transfers: wheelchair-toilet-tub

0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., grab-bars)
2. Independent (or does not require wheelchair)

EFS – Transferring: toilet and tub/shower

#	Description and FIM/SCIM scores
0	Total assistance FIM (both toilet and tub/shower transfers)=0, 1 SCIM=0
1	Partial assistance FIM (toilet and tub/shower transfers)=all other combinations of FIM toilet and tub/shower transfers SCIM=1
2	Independent w/o equipment FIM (both toilet and tub/shower transfers)=7 SCIM=2

EFS – Transfers: toilet

HELPER
0. Activity Does Not Occur—The patient does not transfer on or off the toilet/commode, and is not transferred on or off the toilet/commode by a helper or lifting device. For example, the patient uses only a bedpan and/or urinal. 1. Total Assistance—The patient performs less than 25% of transferring tasks.
2. Maximal Assistance—The patient performs 25 to 49% of transferring tasks.
3. Moderate Assistance—The patient requires more help than touching or performs 50 to 74% of transferring tasks.
4. Minimal Contact Assistance—The patient requires no more help than touching and performs 75% or more of transferring tasks.
5. Supervision or Setup—The patient requires supervision (e.g., standing by, cuing, or coaxing) or setup (positioning sliding board, moving foot rests, etc.).
NO HELPER
6. Modified Independence—The patient requires an adaptive or assistive device such as a sliding board, a lift, grab bars, bedside commode, or special seat; or takes more than a reasonable amount of time to complete the activity; or there are safety considerations. In this case, a prosthesis or orthosis is considered an assistive device if used for the transfer.
7. Complete Independence- *If walking*, patient approaches, sits down on a standard toilet, and gets up from a standard toilet. The patient performs the activity safely. *If in a wheelchair*, patient approaches toilet, locks brakes, lifts foot rests, removes arm rests if necessary, and does either a standing pivot or sliding transfer (without a board) and returns. The patient performs the activity safely.

SCIM – Transfers: wheelchair-toilet-tub

0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., grab-bars)
2. Independent (or does not require wheelchair)

EFS

#	Description and FIM/SCIM scores
	See previous descriptor for transferring: toilet and tub/shower. Also note that FIM tub/shower transfer is one item. The more frequent method is coded. If tub and shower are equally used, the lower score is recorded.

EFS – Transfers: tub

HELPER
0. If the patient does NOT transfer into and out of a tub OR shower, code Transfers: Tub as "0," and leave Transfers: Shower blank. Code "0" may be used for Transfers: Tub on admission and discharge. 1. Total Assistance—The patient performs less than 25% of transferring tasks.
2. Maximal Assistance—The patient performs 25 to 49% of transferring tasks.
3. Moderate Assistance—The patient requires more help than touching or performs 50 to 74% of transferring tasks.
4. Minimal Contact Assistance—The patient requires no more help than touching and performs 75% or more of transferring tasks.
5. Supervision or Setup—The patient requires supervision (e.g., standing by, cuing, or coaxing) or setup (positioning sliding board, moving foot rests, etc.).
NO HELPER
6. Modified Independence—The patient requires an adaptive or assistive device (including a prosthesis or orthosis) such as a sliding board, a lift, grab bars, or special seat; or takes more than a reasonable amount of time to complete the activity; or there are safety considerations.
7. Complete Independence - *If walking*, the patient approaches a tub, and gets into and out of it. The patient performs the activity safely. *If in a wheelchair*, the patient approaches a tub, locks brakes, lifts foot rests, removes arm rests if necessary, and does either a standing pivot or sliding transfer (without a board) and returns. The patient performs the activity safely.

SCIM – Transfers: wheelchair-toilet-tub

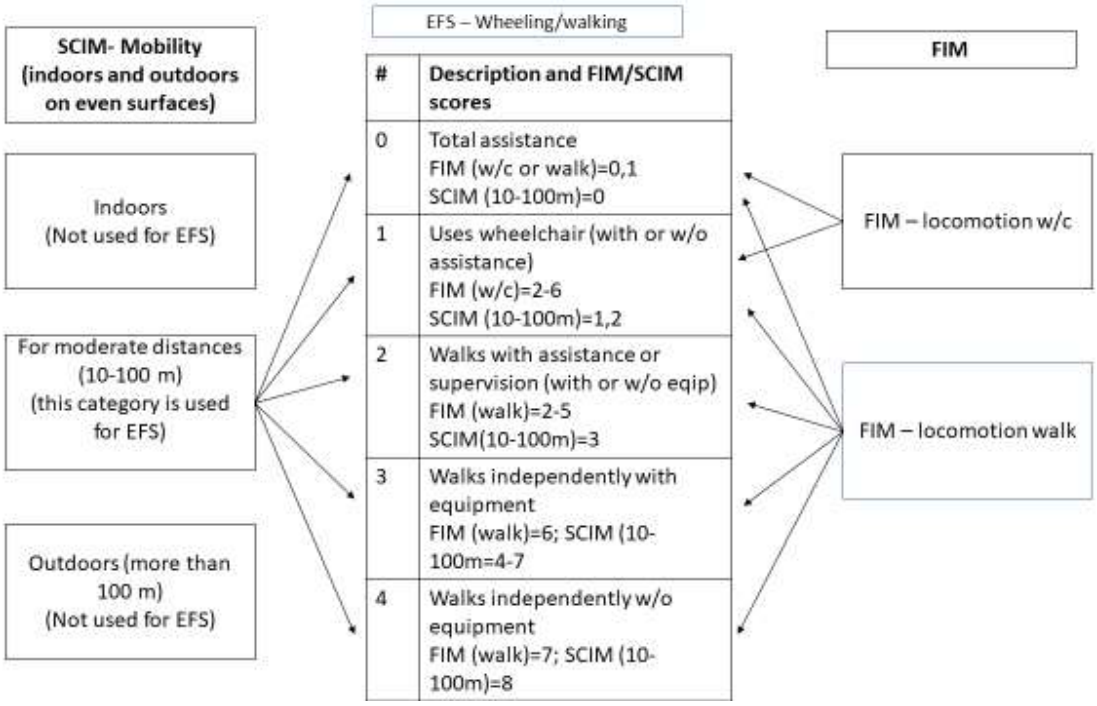
0. Requires total assistance
1. Needs partial assistance and/or supervision, and/or adaptive devices (e.g., grab-bars)
2. Independent (or does not require wheelchair)

EFS

#	Description and FIM/SCIM scores
	See previous descriptor for transferring: toilet and tub/shower. Also note that FIM tub/shower transfer is one item. The more frequent method is coded. If tub and shower are equally used, the lower score is recorded.

EFS – Transfers: shower

HELPER
0. If the patient does NOT transfer into and out of a tub OR shower, code Tub Transfer as "0," and leave Shower Transfer blank. Do not use code "0" for Shower Transfer. 1. Total Assistance—The patient performs less than 25% of transferring tasks.
2. Maximal Assistance—The patient performs 25 to 49% of transferring tasks.
3. Moderate Assistance—The patient requires more help than touching or performs 50 to 74% of transferring tasks.
4. Minimal Contact Assistance—The patient requires no more help than touching and performs 75% or more of transferring tasks.
5. Supervision or Setup—The patient requires supervision (e.g., standing by, cuing, or coaxing) or setup (positioning sliding board, moving foot rests, etc.).
NO HELPER
6. Modified Independence—The patient requires an adaptive or assistive device (including a prosthesis or orthosis) such as a sliding board, a lift, grab bars, or special seat; or takes more than a reasonable amount of time to complete the activity; or there are safety considerations.
7. Complete Independence - *If walking*, the patient approaches a shower stall, and gets into and out of it. The patient performs the activity safely. *If in a wheelchair*, the patient approaches a shower stall, locks brakes, lifts foot rests, removes arm rests if necessary, and does either a standing pivot or sliding transfer (without a board) and returns. The patient performs the activity safely.



SCIM – Mobility (indoors and outdoors, on even surface): Indoors

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal walking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

EFS – Wheeling/walking

#	Description and FIM/SCIM scores
0	See descriptor for EFS wheeling/walking – do not use SCIM – Mobility (indoors and outdoors, on even surface): Indoors – inadequate FIM equivalent

FIM – Locomotion w/c

HELPER
0. The subject does not use a wheelchair and is not pushed in a wheelchair by a helper. 1. Total Assistance- The patient performs less than 25% of effort, or requires the assistance of two people, or wheels less than 50 ft. (15 meters).
2. Maximal Assistance- The patient performs 25-49% of locomotion effort to go a minimum of 50 feet (15 meters) and requires the assistance of one person only.
3. Moderate Assistance- The patient performs 50-74% of locomotion effort to go a minimum of 150 feet (50 meters).
4. Minimal Contact Assistance- The patient performs 75% or more of locomotion effort to go a minimum of 150 feet (50 meters).
5. Supervision - The patient requires standby supervision, cuing, or coaxing to go a minimum of 150 feet (50 meters) in a wheelchair. (Coded the same as 5 with no helper – below)
NO HELPER
5. Exception (Household Locomotion) – The patient operates a manual or motorized wheelchair *independently* only short distances (a minimum of 50 feet or 15 meters). (coded the same as five above with helper)
6. Modified Independence- The patient operates a manual or motorized wheelchair independently for a minimum of 150 feet (50 meters); turns around; maneuvers the chair to a table, bed, toilet; negotiates at least a 3% grade; and maneuvers on rugs and over door sills.
7. This score is not used if patient uses a w/c for locomotion.

SCIM – Mobility (indoors and outdoors, on even surface): For moderate distances (10-100 m)

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal walking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

EFS – Wheeling/walking

#	Description and FIM/SCIM scores
0	Total assistance FIM (w/c or walk)=0,1 SCIM (10-100m)=0
1	Uses wheelchair (with or w/o assistance) FIM (w/c)=2-6 SCIM (10-100m)=1,2
2	Walks with assistance or supervision (with or w/o eqip) FIM (walk)=2-5 SCIM(10-100m)=3
3	Walks independently with equipment FIM (walk)=6; SCIM (10-100m=4-7
4	Walks independently w/o equipment FIM (walk)=7; SCIM (10-100m)=8

FIM – Locomotion w/c

HELPER
0. The subject does not use a wheelchair and is not pushed in a wheelchair by a helper. 1. Total Assistance- The patient performs less than 25% of effort, or requires the assistance of two people, or wheels less than 50 ft. (15 meters).
2. Maximal Assistance- The patient performs 25-49% of locomotion effort to go a minimum of 50 feet (15 meters) and requires the assistance of one person only.
3. Moderate Assistance- The patient performs 50-74% of locomotion effort to go a minimum of 150 feet (50 meters).
4. Minimal Contact Assistance- The patient performs 75% or more of locomotion effort to go a minimum of 150 feet (50 meters).
5. Supervision - The patient requires standby supervision, cuing, or coaxing to go a minimum of 150 feet (50 meters) in a wheelchair. (Coded the same as 5 with no helper – below)
NO HELPER
5. Exception (Household Locomotion) – The patient operates a manual or motorized wheelchair independently only short distances (a minimum of 50 feet or 15 meters). (coded the same as five above with helper)
6. Modified Independence- The patient operates a manual or motorized wheelchair independently for a minimum of 150 feet (50 meters); turns around; maneuvers the chair to a table, bed, toilet; negotiates at least a 3% grade; and maneuvers on rugs and over door sills.
7. This score is not used if patient uses a w/c for locomotion.

SCIM – Mobility (indoors and outdoors, on even surface):Outdoors (more than 100 m)

0. Requires total assistance
1. Needs electric wheelchair or partial assistance to operate manual wheelchair
2. Moves independently in manual wheelchair
3. Requires supervision while walking (with or without devices)
4. Walks with a walking frame or crutches (swing)
5. Walks with crutches or two canes (reciprocal walking)
6. Walks with one cane
7. Needs leg orthosis only
8. Walks without walking aids

EFS – Wheeling/walking

#	Description and FIM/SCIM scores
0	See previous descriptor for wheeling/walking – do not use SCIM – Mobility (indoors and outdoors, on even surface): Outdoors (more than 100m) – no FIM distance equivalent

FIM – Locomotion w/c

HELPER
0. The subject does not use a wheelchair and is not pushed in a wheelchair by a helper. 1. Total Assistance- The patient performs less than 25% of effort, or requires the assistance of two people, or wheels less than 50 ft. (15 meters).
2. Maximal Assistance- The patient performs 25-49% of locomotion effort to go a minimum of 50 feet (15 meters) and requires the assistance of one person only.
3. Moderate Assistance- The patient performs 50-74% of locomotion effort to go a minimum of 150 feet (50 meters).
4. Minimal Contact Assistance- The patient performs 75% or more of locomotion effort to go a minimum of 150 feet (50 meters).
5. Supervision -The patient requires standby supervision, cuing, or coaxing to go a minimum of 150 feet (50 meters) in a wheelchair. (Coded the same as 5 with no helper – below)
NO HELPER
5. Exception (Household Locomotion)—The patient operates a manual or motorized wheelchair independently only short distances (a minimum of 50 feet or 15 meters). (coded the same as five above with helper)
6. Modified Independence- The patient operates a manual or motorized wheelchair independently for a minimum of 150 feet (50 meters); turns around; maneuvers the chair to a table, bed, toilet; negotiates at least a 3% grade; and maneuvers on rugs and over door sills.
7. This score is not used if patient uses a w/c for locomotion.

SCIM – Stair Management

0. Unable to ascend or descend stairs
1. Ascends and descends at least 3 steps with support or supervision of another person
2. Ascends and descends at least 3 steps with support of handrail and/or crutch cane
3. Ascends and descends at least 3 steps without any support or supervision

EFS - Climbing Stairs

0	Does not do or total assist SCIM=0 FIM=0,1
1	Partial Assistance SCIM=1 FIM=2-5
2	Independent with equipment SCIM=2 FIM=6
3	Independent w/o equipment SCIM=3 FIM=7

FIM – Locomotion: Stairs

HELPER
0. Activity Does Not Occur- The subject does not go up or down stairs, and a helper does not carry the subject up or down stairs. 1.Total Assistance - The patient performs less than 25% of the effort, or requires the assistance of two people, or goes up and down fewer than 4 stairs.
2. Maximal Assistance - The patient performs 25-49% of the effort to go up or down 4 to 6 stairs, and requires assistance of one person only.
3. Moderate Assistance- The patient performs 50-74% of the effort to go up or down one flight of stairs.
4. Minimal Contact Assistance- The patient performs 75% or more of the effort to go up or down one flight of stairs.
5. Supervision- The patient requires supervision (standing by, cuing, or coaxing) to go a go up or down one flight of stairs. (coded the same as 5 with no helper – below)
NO HELPER
5.. Exception (Household Ambulation)—The patient goes up and down 4 to 6 stairs *independently*, with or without a device. The activity takes more than a reasonable amount of time, or there are safety considerations. (coded the same as 5 with helper – above)
6. Modified Independence - The patient goes up and down at least one flight of stairs but requires a side support, handrail, cane, or portable supports; or the activity takes more than a reasonable amount of time; or there are safety considerations.
7. Complete Independence- The patient safely goes up and down at least one flight of stairs without depending on any type of handrail or support.

ground-wheelchair

0. Requires assistance
1. Transfers independent with or without adaptive devices (or does not require wheelchair)

EFS - ?

#	Description and FIM/SCIM scores
	Not able to map since no comparable items in FIM

FIM –

Transfers: wheelchair-car

0. Requires total assistance
1. Needs partial assistance and/or supervision and/or adaptive devices
2. Transfers independent; does not require adaptive devices (or does not require wheelchair)

EFS - ?

#	Description and FIM/SCIM scores
	Not able to map since no comparable items in FIM

FIM –
