Creating a common metric based on existing activities of daily living tools to enable standardized reporting of functioning outcomes achieved during rehabilitation

R. Maritz, A. Tennant, C. Fellinghauer, G. Stucki, Birgit Prodinger

Angaben zur Veröffentlichung / Publication details:

Maritz, R., A. Tennant, C. Fellinghauer, G. Stucki, and Birgit Prodinger. 2020. "Creating a common metric based on existing activities of daily living tools to enable standardized reporting of functioning outcomes achieved during rehabilitation." *Journal of Rehabilitation Medicine* 52 (7): 1–9. https://doi.org/10.2340/16501977-2711.

J Rehabil Med 2020; 52: jrm00085

ORIGINAL REPORT

CREATING A COMMON METRIC BASED ON EXISTING ACTIVITIES OF DAILY LIVING TOOLS TO ENABLE STANDARDIZED REPORTING OF FUNCTIONING OUTCOMES ACHIEVED DURING REHABILITATION

Roxanne MARITZ, MA[1,2], Alan TENNANT, PhD[1,2], Carolina FELLINGHAUER, PhD[1], Gerold STUCKI, MD, MS[1,2] and Birgit PRODINGER, PhD[1,2,3], on behalf of the NRP74 StARS clinics[4]

From the [1]Swiss Paraplegic Research, Nottwil, Switzerland [2]Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland, [3]Faculty of Applied Health and Social Sciences, Technical University of Applied Sciences Rosenheim, Rosenheim, Germany, [4]NRP74 StARS clinics involved in the underlying study (Clinica Hildebrand Centro di riabilitazione Brissago – Giovanni Rabito, Kliniken Valens – Stefan Bachmann, RehaClinic Bad Zurzach and RehaClinic Kilchberg – Serge Altmann and Peter S. Sandor, Klinik Bethesda Tschugg)

Objective: **Many different assessment tools are used to assess functioning in rehabilitation; this limits the comparability and aggregation of respective data. The aim of this study was to outline the development of an International Classification of Functioning, Disability and Health (ICF)-based interval-scaled common metric for 2 assessment tools assessing activities of daily living: the Functional Independence Measure (FIM™) and the Extended Barthel Index (EBI), used in Swiss national rehabilitation quality reports.**

Methods: **The conceptual equivalence of the 2 tools was assessed through their linking to the ICF. The Rasch measurement model was then applied to create a common metric including FIM™ and EBI.**

Subjects: **Secondary analysis of a sample of 265 neurological patients from 5 Swiss clinics.**

Results: **ICF linking found conceptual coherency of the tools. An interval-scaled common metric, including FIM™ and EBI, could be established, given fit to the Rasch model in the related analyses.**

Conclusion: **The ICF-based and interval-scaled common metric enables comparison of patients' and clinics' functioning outcomes when different activities of daily living tools are used. The common metric can be included in a Standardized Assessment and Reporting System for functioning information in order to enable data aggregation and comparability.**

Key words: outcome assessment (healthcare); psychometrics; rehabilitation; activities of daily living; Rasch Measurement Model; Functional Independence Measure; Barthel Index; quality in healthcare.

Accepted Jun 11, 2020; Epub ahead of print Jul 10, 2020

J Rehabil Med 2020; 52: jrm00085

Correspondence address: Roxanne Maritz, Swiss Paraplegic Research, Guido A. Zäch Strasse 4, 6207 Nottwil, Switzerland, E-Mail: roxanne.maritz@paraplegie.ch

LAY ABSTRACT

In our study we developed a common metric serving as a neutral comparator of two assessment tools which are used for assessing activities of daily living in rehabilitation patients in Switzerland. This common metric enables clinicians to use different established assessment tools assessing the same information, while being able to compare the respective information from those different scales on a larger level e.g. comparisons across clinics using different tools. This study is based on the example of Switzerland, where rehabilitation clinics can choose one of two measurement tools assessing activities of daily living, to report their outcome quality. With the common metric, the results from all the clinics can be compared with each other, no matter what tool was used for the assessment, enabling learning and improvement processes.

Functioning is the key indicator for rehabilitation as a health strategy (1). In order to strengthen rehabilitation it is essential to integrate functioning information, through the WHO's International Classification of Functioning, Disability and Health (ICF), into national health information systems, including reports on health-care quality (2–4). Functioning information, including information on activities of daily living (ADL), is often collected using a variety of assessment tools, which can limit comparability across patients and clinics (5). Two prominent examples of such tools are the Functional Independence Measure (FIM™) and the Barthel Index (BI) (6, 7). Noteworthy, of these well-established tools there are many country and rehabilitation group adapted versions, such as the Extended Barthel Index (EBI) (8), the Modified Barthel Index (9) and the United Kingdom Functional Assessment Measure (UK FIM+FAM) (10). While all of these assessment tools focus on ADLs, their items and scoring structures differ.

There are 2 options for enhancing the comparability of functioning information: first, to define a single assessment tool as the standard, and, secondly, to establish a transformation system between existing established assessment tools. The first option would be difficult to achieve, as there are various reasons for the heterogeneity of the assessment tools in use, such as clinical utility and clinic-specific standards. The latter option, in which a common metric for functioning information is developed, to enable comparison and aggregation of information collected with different tools that measure the same concept, is more feasible (5, 11–15). For this purpose, a Standardized Assessment and Reporting

 ISSN 1650-1977 doi: 10.2340/16501977-2711

System (StARS) for functioning information, with an ICF-based interval-scaled neutral common metric as core element, would enable clinicians to continue using different assessment tools, while at the same time enabling the aggregation and comparison of conceptually equivalent information (5, 13). This common metric could serve as a reporting reference, e.g. for national rehabilitation quality reports, allowing for comparisons between institutions using different tools. A common metric also allows for the transformation of the score of one tool into the score of another conceptually equivalent tool. Furthermore, it provides a transformation of the ordinal-scaled ADL scores to an interval-scale level, which is required to calculate means and change scores between admission and discharge (16).

In order to demonstrate how a StARS of functioning information can be established, this paper takes Switzerland's national rehabilitation quality reports as an example, providing a simplified illustration for the heterogeneous landscape of assessment tools. For Swiss national quality reports in musculoskeletal and neurological rehabilitation, clinics can choose to report with the FIM™ or the EBI, but this impedes the comparison of clinics that use different assessment tools.

Earlier research by Prodinger et al. provided a score transformation of the FIM™ 13-items motor score and the BI on the basis of the Rasch model (17). While the FIM™ 13-item motor scale has been studied extensively (18), its 5 items of cognition have received less attention. Nevertheless, cognitive impairment is important in neurological disorders (19). The importance of cognition was the reason for development of the EBI, extending the BI with 6 cognitive items (8). The current paper therefore seeks to build on the evidence established by Prodinger et al. regarding the motor scales of these 2 assessment tools (17), using the same psychometric approach, but extending it to the 2 assessment tools' versions, including cognitive items. Furthermore, this paper provides a concrete example of how an ICF-based and interval-scaled common metric as a core of a StARS (5), can be created, so that the outcomes of different rehabilitation clinics, using different assessment tools assessing ADLs, can be compared and aggregated.

The objective of this study was to create an ICF-based interval-scaled common metric as a core element of a StARS for functioning information, based on the example of Swiss national quality reports. The approach included the assessment of the 2 key requirements for standardized reporting of health information: (*i*) to determine whether the 2 assessment tools can be considered conceptually equivalent, and (*ii*) to examine whether a reference metric including the FIM™ and EBI can be established by applying the Rasch model (20).

METHODS

Setting and subjects

In Switzerland, the National Association for Quality Development in Hospitals and Clinics (ANQ) coordinates the measurement and public reporting of outcome quality indicators for all rehabilitation clinics (21). For this purpose, functioning information from every patient in neurological and musculoskeletal rehabilitation is collected. Clinics can choose 1 of the 2 tools, either the FIM™ (18-item version) or the EBI for this part of the data collection. To overcome the issue of comparability, the ANQ has commissioned the development of an expert-based transformation algorithm of the 2 assessment tools, called the ANQ ADL Score. In order to create and validate the expert-based transformation algorithm, the Institute of Medical Sociology and Rehabilitation Science from the Charité – Universitätsmedizin Berlin, Germany conducted the respective study from 2015 to 2017 (22). The validation sample included 265 patients undergoing neurorehabilitation from 5 Swiss rehabilitation clinics, representative of the whole continuum of score ranges of the 2 assessment tools. All patients were assessed at admission with both assessment tools, effectively providing a reliable basis for scale-equating procedures, i.e. a common person design (23, 24). Both tools were assessed either in German (4 clinics) or Italian (1 clinic). The data collected in the ANQ ADL Score study, i.e. a prior research project, were used in secondary data analysis in the current study to provide the basis for the ICF-based and interval-scaled common metric. Both the ANQ ADL Score study and the current study were given ethics approval by the respective Swiss Ethics Commission.

ADL assessment tools

The FIM™ is an assessment tool administered by health professionals comprising 18 items. In order to qualify for FIM™ administration, the health professionals received training provided by the ANQ according to the respective FIM™ policy. FIM™ consists of 13 motor items and 5 cognitive items. All items are scored from 1 (total assistance) to 7 (complete independence), summing to a total score ranging between 18 and 126 (25).

The EBI is an assessment tool with 16 items, administered by health professionals. A user manual for the administration of EBI is available, but there is no specific training. Ten motor items are based on the original BI (26), 6 items cover cognitive functioning aspects, of which 5 are derived from the cognitive FIM™ items. One cognitive item is unique to EBI and refers to vision and neglect. All items are scored from 0 to 4, resulting in a total score of 0–64 (8). Not all EBI items contain all scoring categories (e.g. Item 1 Feeding can be scored 0, 2, 3 or 4), therefore an adapted 0–50 scoring version was proposed for Rasch analyses with EBI, which was taken as a basis for this study (27). A conversion from EBI 0–64 scores to EBI 0–50 scores, referred to as EBI50, on item basis can be found in Appendix S1[1].

Recent studies for both tools showed, that in the context of national quality reports they measure a unidimensional construct and can be reported as total scores on the interval-level, when Rasch-based transformation with bi-factor equivalent design is applied (27, 28). In these studies, neither tool showed differential item functioning for sex, age, nationality, healthcare insurance status of patients, time-point of measurement, rehabilitation group (neurological or musculoskeletal) or clinic language (German, French or Italian).

[1]https://doi.org/10.2340/16501977-2711

ICF linking

The first part of data analysis entailed linking each item of the respective tools to the ICF by using the ICF linking rules, an established method to enable comparability of health information (29). In order to satisfy the first requirement for standardized reporting and scale equating, i.e. to demonstrate the conceptual equivalence of the 2 assessment tools to be integrated into the common metric, the items from each assessment tool were linked to corresponding ICF categories. Furthermore, the perspectives from which information is collected and the categorization of response options were identified for both tools, in accordance with the ICF linking rules. Two researchers, with extensive linking experience (Maritz R., Selb M.), independently linked all items of the FIM™ and the EBI to the ICF, following step by step the refined version of the ICF linking rules (29) in Microsoft® Excel. The results of the independent linking were then compared and discussed. When no agreement on linking of an item could be reached, a third researcher (Prodinger B.) was involved to give advice and reach agreement.

Rasch analysis

The second part of the data analysis was based on Rasch analysis. In order to satisfy the second requirement for standardized reporting, i.e. score equivalence, the polytomous partial credit Rasch measurement model and associated requirements for equating of instruments were applied to derive an interval-scaled common metric from ordinal data (23, 30). These requirements include: unidimensionality; item invariance across sample subgroups, such as age or sex; and local independence, i.e. the demonstration that responses to any item should depend only on the trait (functional independence in the case of EBI and FIM™) and not on responses to other items (31). The analyses were conducted on the total score level of the 2 assessment tools. This was based on the reasoning that: (*i*) the total scores reflect the level of reporting in the national quality reports; (*ii*) previous findings support that the 2 tools can be reported as unidimensional metrics on the total score level (27, 28), representing the construct of functional independence, even though they incorporate both motor and cognitive items; and (*iii*) the recommendations for scale equating by Andrich are fullfilled (23). The data from the validation sample of the ANQ ADL Score study (described above under the subheading "Setting and subjects") were used for Rasch analyses. Basic sample characteristics and descriptive statistics were conducted using Stata Version 14.2, Rasch analysis was conducted with RUMM2030 professional version 5.4.

The analytical focus gave reference to the following 6 key criteria, helping to judge if fit to the Rasch model and its requirements for equating of 2 instruments was achieved: (*i*) the class-interval based conditional test of fit, assessing the observed and the expected scores under the model conditional on each total score through a Pearson χ^2 test; (*ii*) the item-trait interaction χ^2 test, reflecting fit of the data to the Rasch model, also referring to observed and expected scores on the level of class-intervals; (*iii*) the reliability indexes, reported as Cronbach's alpha and person separation index (PSI); (*iv*) differential item functioning (DIF), indicating if there is invariance for different subgroups; (*v*) unidimensionality, expressed as percentage of significant *t*-tests, using individual *t*-tests comparing person-ability estimates for each respondent derived from the subtest analysis; and (*vi*) threshold ordering, indicating whether the different scoring categories of an assessment scale are represented in a successive order (16, 23, 30). Acceptable levels of the key criteria are represented in the bottom row of the corresponding results Table I.

In order to fully examine the defined key criteria, a 2-tiered analysis was used to deal with some restrictions of the analysis software, as FIM™ has 109 scoring options (ranging from 18 to 126) and the RUMM2030 software allows for inclusion of only 101 scoring options. First, analysis of the assessment tools' total scores was performed, in which the FIM™ total scores were rescaled to 0 to 100 together with the EBI50 total scores. This first step is shown in the first row of the corresponding result in Table I, allowing for a conditional test of fit in which both total scores served as 2 items. The FIM™ score was then re-weighted by 1.09 to give the usual operational score of the FIM™. Secondly, the FIM™ items were divided into 2 testlets and the EBI50 items were combined in a third testlet. Which FIM™ item was contained in which testlet, was based on a previous research project about FIM™ total scores for use in national quality reports (28) and is indicated in the legend of Table I. This second analysis step enabled values of the variance in the latent estimate to be obtained, which imply the degree of local dependency remaining in the testlets (shown in the second row of the corresponding result in Table I). Furthermore, the 2 FIM™ testlets could then be taken together as a single super-testlet in a paired *t*-test analysis with the EBI total score, thus addressing the software limitation described, so that the FIM™ total scores can be reported on their original range.

Differential item functioning strategy

DIF was tested for sex, age (4 groups based on interquartile ranges), healthcare insurance status (general, semi-private, private), nationality (Swiss/other), duration of rehabilitation (4 groups based interquartile ranges) and clinic (5 rehabilitation clinics involved) using 4 class intervals. When uniform or non-uniform DIF *p*-values were <0.05, DIF was considered to be present, and the respective testlets were split, starting with the highest DIF, and continuing until no further DIF was present (32). The split and unsplit solutions were then contrasted on the basis of the Rasch person estimates, anchored to each other with an unsplit testlet free of DIF. To determine whether DIF split was necessary for the transformation table, an effect size calculation was performed, based on a Cohen's D calculation, including the mean of the person estimates, their standard deviations, and the sample size of the split and unsplit version (32). If the resulting

Table I. Results of the FIM™ EBI Rasch equating analyses

Data basis	Conditional test of fit *p*-value (DF)	PSI	α	Item- trait χ^2 *p*-value (DF)	Threshold disordering	DIF (testlets)	A Variance	paired *t*-test	Comment
FIM™0–100 – EBI50 total scores	0.504 (112)	0.958	0.848	0.861 (6)	No disordering	Clinic (FIM™, EBI50)	Not applicable	2.64 %	
FIM™ – EBI50 3 testlets FIM1, FIM2, EBI50	Not applicable	0.973	0.981	0.921 (9)	No disordering	Clinic (FIM1, FIM2, EI50)	0.999	6.04 % (lower CI= 3.4%)	FIM testlets taken together for *t*-test and transformation table
Acceptable values	>0.05	>0.70	>0.70	>0.05	No disordering	No substantial DIF	>0.90	<5.00 % (at least lower CI)	

PSI: person separation index; DIF: differential item functioning; A Variance: explained common variance (only available for testlets); DF: degrees of freedom; CI: 95% confidence interval. The FIM1 testlet included FIM™ items A, C, E, G, I, K, M, O & Q and the FIM2 testlet included FIM™ items B, D, F, H J, L, N, P & R. EBI: Extended Barthel Index; FIM™: Functional Independence Measure.

effect size was below 0.1, DIF was considered small and no action was taken, i.e. the unsplit solution was retained (33).

Common metric

If model fit was achieved with both Rasch analysis steps, an interval-scaled common metric was created based on the Rasch location estimates. The common metric is based on a paired t-test from the second Rasch analysis step, with the EBI total score testlet on the one hand and both FIM(TM) testlets together on the other hand. Likewise, the FIM(TM) total scores can be reported on their original range (18–126) and are therefore considered more accurate. The common metric was designed to range from 0 (complete dependence in ADL) to 100 (complete independence in ADL), oriented at similar research projects (13, 17), reflecting minimum and maximum logit estimates derived from the joint analysis.

RESULTS

International Classification of Functioning, Disability and Health linking

An overview of the ICF linking is shown in Table II, and detailed linking at the level of the items is shown in Appendix S2[1]. Both assessment tools' items represent a dependency perspective and all items responses were categorized in the form of an intensity of this dependency (1–7 for FIM™, 0–4 for EBI). The content of the FIM™ items was reflected in 24 ICF categories, and the content of the EBI in 26 ICF categories. Both assessment tools cover the same ICF categories, with the exception of EBI item 16 vision/neglect, linked to the ICF categories

Table II. Overview of the International Classification of Functioning, Disability and Health (ICF) linking table for the assessment tools: FIM™ and EBI

Perspective of items Categorization of item responses		FIM™ Dependency Intensity	EBI Dependency Intensity
ICF Code & Label			
b	BODY FUNCTIONS		
b1	Mental Functions		
b144	Memory functions	18) Memory	15) Memory
b156	Perceptual functions		16) Vision/Neglect
b2	Sensory functions and pain		
b210	Seeing functions		16) Vision/Neglect
b5	Functions of the digestive, metabolic and endocrine systems		
b525	Defecation functions	08) Bowel management	09) Bowel control
b6	Genitourinary and reproductive functions		
b620	Urination functions	07) Bladder management	10) Bladder control
b7	Neuromusculoskeletal and movement related functions		
d	ACTIVITIES AND PARTICIPATION		
d1	Learning and applying knowledge		
d175	Solving problems	17) Problem solving	14) Problem solving
d2	General tasks and demands		
d3	Communication		
d310	Communicating with - receiving - spoken messages	14) Comprehension	11) Comprehension
d315	Communicating with - receiving - nonverbal messages	14) Comprehension	11) Comprehension
d320	Communicating with - receiving - formal sign language messages	14) Comprehension	11) Comprehension
d325	Communicating with - receiving - written messages	14) Comprehension	11) Comprehension
d330	Speaking	15) Expression	12) Expression
d335	Producing nonverbal messages	15) Expression	12) Expression
d340	Producing messages in formal sign language	15) Expression	12) Expression
d345	Writing messages	15) Expression	12) Expression
d4	Mobility		
d410	Changing basic body position	09) Transfer bed-chair-wheelchair 10) Transfer toilet 11) Transfer tub/shower	05) Transfers
d420	Transferring oneself	09) Transfer bed-chair-wheelchair 10) Transfer toilet 11) Transfer tub/shower	05) Transfers
d450	Walking	12) Walking/using wheelchair	06) Mobility
d455	Moving around	13) Stairs	07) Stairs
d465	Moving around using equipment	12) Walking/using wheelchair	06) Mobility
d5	Self-care		
d510	Washing oneself	2) Grooming 3) Bathing	02) Grooming 04) Bathing
d520	Caring for body parts	2) Grooming	02) Grooming
d530	Toileting	6) Toileting 7) Bladder management 8) Bowel management	08) Toilet use 09) Bowel control 10) Bladder control
d540	Dressing	4) Dressing upper body 5) Dressing lower body	03) Dressing
d550	Eating	1) Eating	01) Feeding
d560	Drinking	1) Eating	01) Feeding
d7	Interpersonal interactions and relationships		
d710	Basic interpersonal relationships	16) Social interaction	13) Social interaction

EBI: Extended Barthel Index, FIM™: Functional Independence Measure.

b156 Perceptual functions and b210 Seeing functions that are not reflected in the FIM™. Both tools covered predominantly the activities and participation categories of the ICF, with a focus on the ICF chapters *d3 Communication, d4 Mobility and d5 Self-care*. Due to the high level of concordance, the tools were considered conceptually equivalent, i.e. measuring the same latent trait, both covering content related to the concept of functioning.

Sample characteristics

The sample incorporated 265 patients in neurological rehabilitation in Switzerland from 2016 and 2017. Four clinics from the German-speaking and one clinic from the Italian-speaking region of Switzerland each provided between 31 and 70 cases. The sex distribution was 50.8% male ($n=123$) and 49.2% female ($n=119$). The mean age of the sample was 67.2 years, ranging from 18 to 92 years, while the mean age of the cases of the different clinics ranged from 62.9 to 73.9 years. The majority (90.0%) were Swiss citizens, 10.0% had other nationalities. Almost three-quarters (74.0%) had general healthcare insurance, 18.6% semi-private insurance and 7.4% private insurance. The mean rehabilitation duration time was 37.7 days, ranging from 5 to 150 days (standard deviation 25.4). Minimal attained FIM™ score of 18 was attained by 7 cases, the maximal score of 126 was not achieved. The minimal EBI score of 0 was achieved by 2, the maximal score by 11 cases. There were missing values for the variables sex, age, healthcare insurance status, rehabilitation duration (each $n=23$), and origin ($n=24$).

Rasch analysis

Results of the Rasch equating analyses are shown in Table I. Fit of the data as total scores (first analysis step, first line Table I), and as 3 testlets (second analysis step with 1 EBI testlet, 2 FIM™ super-items, second line of Table I) showed good fit to the Rasch model. The conditional test of fit in the total score analysis and the item-trait χ^2 in both analyses were all non-significant. Reliability was consistent with high-stakes clinical decision-making. All the common variance was subsequently included in the interval-scaled transformation, and both analyses satisfied the monotonic relationship with the functioning trait, thus showing no disordering of thresholds, again a requirement for successful test equating at the scale score level. DIF was present at the clinic level for both tools. A comparison of person estimates between a non-split and split- solution resulted in an effect size of 0.02; thus, it was considered marginal and no action was taken. Detailed information on the DIF strategy and the effect size calculation are shown in Appendix S3[1].

Common metric

Based on the second analysis step, a common metric was created (see Table III), using the total scores of FIM™ and EBI as testlets. The related transformation table for FIM™ and EBI50 total raw scores can be retrieved from the metric and is shown in Appendix S4[1]. For example a FIM™ raw score of 114 can be translated to 72.9 in the common metric, or an EBI50 raw score of 47. The common metric also shows that the operational range of the FIM™ (common metric values from 8.4 to 96.6) is larger than that of the EBI (common metric values from 9.3 to 81.1), indicating that the FIM™ covers a wider range of patient abilities and that there might be ceiling effects with the EBI. The operational ranges of the 2 assessment tools, in contrast to the common metric, are shown in Fig. 1. The fact that the operational range of the common metric is wider than the range of both assessment tools on its own, can be explained by the fact that the calculation of the common metric is based on patients assessed with both tools together, i.e. the common person design of the sample.

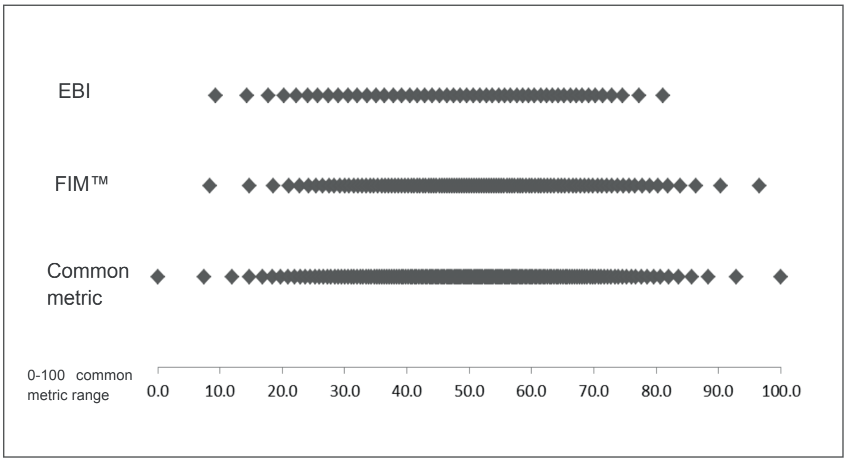

Fig. 1. The 0–100 ICF-based and interval-scaled common metric including the Extended Barthel Index (EBI) and the Functional Independence Measure (FIM™). ICF: International Classification of Functioning, Disability and Health.

Table III. International Classification of Functioning, Disability and Health (ICF)-based interval-scaled common metric 0–100, including Functional Independence Measure (FIM™) and Extended Barthel Index (EBI)50

ICF-based interval-scaled common metric 0-100 including FIM™ and EBI50

common metric raw	RE common metric	Common metric 0-100
0	-3,618	0,0
1	-3,089	7,4
2	-2,763	12,0
3	-2,563	14,8
4	-2,419	16,8
5	-2,304	18,4
6	-2,208	19,8
7	-2,125	20,9
8	-2,052	21,9
9	-1,985	22,9
10	-1,924	23,7
11	-1,868	24,5
12	-1,815	25,3
13	-1,765	26,0
14	-1,718	26,6
15	-1,673	27,2
16	-1,63	27,9
17	-1,588	28,4
18	-1,548	29,0
19	-1,509	29,5
20	-1,472	30,1
21	-1,435	30,6
22	-1,399	31,1
23	-1,364	31,6
24	-1,33	32,1
25	-1,296	32,5
26	-1,263	33,0
27	-1,23	33,5
28	-1,199	33,9
29	-1,167	34,3
30	-1,136	34,8
31	-1,106	35,2
32	-1,076	35,6
33	-1,046	36,0
34	-1,017	36,4
35	-0,988	36,8
36	-0,96	37,2
37	-0,932	37,6
38	-0,904	38,0
39	-0,877	38,4
40	-0,85	38,8
41	-0,824	39,1
42	-0,797	39,5
43	-0,771	39,9
44	-0,746	40,2
45	-0,72	40,6
46	-0,695	40,9
47	-0,67	41,3
48	-0,646	41,6
49	-0,622	42,0
50	-0,598	42,3
51	-0,574	42,6
52	-0,55	43,0
53	-0,527	43,3
54	-0,504	43,6
55	-0,481	43,9
56	-0,459	44,3
57	-0,436	44,6
58	-0,414	44,9
59	-0,392	45,2
60	-0,37	45,5
61	-0,348	45,8
62	-0,326	46,1
63	-0,305	46,4
64	-0,284	46,7
65	-0,262	47,0
66	-0,241	47,3
67	-0,22	47,6
68	-0,199	47,9
69	-0,178	48,2
70	-0,158	48,5
71	-0,137	48,8
72	-0,116	49,1
73	-0,096	49,3
74	-0,075	49,6
75	-0,055	49,9
76	-0,034	50,2
77	-0,014	50,5
78	0,006	50,8
79	0,027	51,1

common metric raw cont'd	RE common metric	Common metric 0-100
80	0,047	51,3
81	0,067	51,6
82	0,088	51,9
83	0,108	52,2
84	0,128	52,5
85	0,149	52,8
86	0,169	53,1
87	0,19	53,3
88	0,21	53,6
89	0,231	53,9
90	0,251	54,2
91	0,272	54,5
92	0,293	54,8
93	0,313	55,1
94	0,334	55,4
95	0,355	55,7
96	0,376	56,0
97	0,397	56,2
98	0,419	56,6
99	0,44	56,9
100	0,462	57,2
101	0,483	57,5
102	0,505	57,8
103	0,527	58,1
104	0,549	58,4
105	0,571	58,7
106	0,593	59,0
107	0,616	59,3
108	0,638	59,6
109	0,661	59,9
110	0,684	60,3
111	0,707	60,6
112	0,731	60,9
113	0,754	61,2
114	0,778	61,6
115	0,802	61,9
116	0,826	62,3
117	0,85	62,6
118	0,875	62,9
119	0,9	63,3
120	0,925	63,6
121	0,95	64,0
122	0,975	64,3
123	1,001	64,7
124	1,027	65,1
125	1,053	65,4
126	1,08	65,8
127	1,107	66,2
128	1,134	66,6
129	1,162	67,0
130	1,19	67,4
131	1,218	67,8
132	1,247	68,2
133	1,276	68,6
134	1,306	69,0
135	1,337	69,4
136	1,369	69,9
137	1,401	70,3
138	1,434	70,8
139	1,468	71,3
140	1,504	71,8
141	1,541	72,3
142	1,58	72,8
143	1,62	73,4
144	1,663	74,0
145	1,709	74,6
146	1,757	75,3
147	1,809	76,0
148	1,864	76,8
149	1,925	77,7
150	1,991	78,6
151	2,064	79,6
152	2,146	80,8
153	2,241	82,1
154	2,354	83,7
155	2,497	85,7
156	2,693	88,4
157	3,009	92,8
158	3,52	100,0

FIM™ raw	RE FIM™	Common metric 0-100
18	-3,015	8,4
19	-2,568	14,7
20	-2,290	18,6
21	-2,116	21,0
22	-1,989	22,8
23	-1,886	24,3
24	-1,799	25,5
25	-1,723	26,5
26	-1,654	27,5
27	-1,590	28,4
28	-1,531	29,2
29	-1,475	30,0
30	-1,421	30,8
31	-1,370	31,5
32	-1,320	32,2
33	-1,272	32,9
34	-1,226	33,5
35	-1,181	34,1
36	-1,137	34,8
37	-1,094	35,4
38	-1,051	36,0
39	-1,010	36,5
40	-0,970	37,1
41	-0,931	37,6
42	-0,892	38,2
43	-0,854	38,7
44	-0,817	39,2
45	-0,781	39,7
46	-0,745	40,2
47	-0,710	40,7
48	-0,676	41,2
49	-0,642	41,7
50	-0,609	42,2
51	-0,576	42,6
52	-0,544	43,1
53	-0,513	43,5
54	-0,481	43,9
55	-0,450	44,4
56	-0,420	44,8
57	-0,390	45,2
58	-0,360	45,6
59	-0,331	46,0
60	-0,302	46,5
61	-0,273	46,9
62	-0,244	47,3
63	-0,215	47,7
64	-0,187	48,1
65	-0,159	48,5
66	-0,130	48,9
67	-0,102	49,3
68	-0,074	49,6
69	-0,046	50,0
70	-0,018	50,4
71	0,010	50,8
72	0,037	51,2

FIM™ raw cont'd	RE FIM™	Common metric 0-100
73	0,065	51,6
74	0,094	52,0
75	0,122	52,4
76	0,150	52,8
77	0,178	53,2
78	0,207	53,6
79	0,236	54,0
80	0,265	54,4
81	0,294	54,8
82	0,323	55,2
83	0,353	55,6
84	0,383	56,1
85	0,413	56,5
86	0,444	56,9
87	0,475	57,3
88	0,507	57,8
89	0,539	58,2
90	0,571	58,7
91	0,604	59,1
92	0,638	59,6
93	0,672	60,1
94	0,707	60,6
95	0,742	61,1
96	0,778	61,6
97	0,815	62,1
98	0,852	62,6
99	0,891	63,2
100	0,930	63,7
101	0,969	64,3
102	1,010	64,8
103	1,052	65,4
104	1,094	66,0
105	1,138	66,6
106	1,182	67,2
107	1,228	67,9
108	1,275	68,5
109	1,323	69,2
110	1,372	69,9
111	1,423	70,6
112	1,476	71,4
113	1,532	72,1
114	1,589	72,9
115	1,650	73,8
116	1,714	74,7
117	1,782	75,7
118	1,855	76,7
119	1,934	77,8
120	2,020	79,0
121	2,117	80,3
122	2,230	81,9
123	2,367	83,8
124	2,547	86,4
125	2,832	90,4
126	3,280	96,6

EBI50 raw	RE EBI50	Common metric 0-100
0	-2,951	9,3
1	-2,597	14,3
2	-2,349	17,8
3	-2,174	20,2
4	-2,029	22,3
5	-1,900	24,1
6	-1,779	25,8
7	-1,662	27,4
8	-1,548	29,0
9	-1,435	30,6
10	-1,326	32,1
11	-1,219	33,6
12	-1,116	35,1
13	-1,015	36,5
14	-0,917	37,8
15	-0,822	39,2
16	-0,730	40,5
17	-0,640	41,7
18	-0,553	42,9
19	-0,468	44,1
20	-0,385	45,3
21	-0,304	46,4
22	-0,224	47,5
23	-0,147	48,6
24	-0,071	49,7
25	0,004	50,7
26	0,077	51,8
27	0,149	52,8
28	0,220	53,8
29	0,290	54,7
30	0,360	55,7
31	0,429	56,7
32	0,497	57,6
33	0,566	58,6
34	0,634	59,6
35	0,702	60,5
36	0,770	61,5
37	0,838	62,4
38	0,906	63,4
39	0,974	64,3
40	1,042	65,3
41	1,111	66,3
42	1,180	67,2
43	1,250	68,2
44	1,324	69,2
45	1,401	70,3
46	1,487	71,5
47	1,588	72,9
48	1,713	74,7
49	1,898	77,3
50	2,173	81,1

RE: Rasch estimate (location); cont'd: continued.

DISCUSSION

This study provides an ICF-based and interval-scaled common metric, including the transformation between FIM™ and EBI for use in national rehabilitation quality reports, and thus facilitates a Standardized Assessment and Reporting System (StARS) for functioning information. In order to create the common metric, 2 steps were followed; first, ICF linking showed that the 2 assessment tools can be considered as conceptually equivalent. In the second step a common metric including the 2 tools could be established, as fit to the Rasch models' requirements for the equating of instruments was achieved. The current study provides an example of how ADL scales assessing the same information can be made comparable through the ICF on the conceptual level and an interval-scaled common reference metric on the scale level. The methodology applied can inform further research, in which conceptually similar functioning information is collected with different tools and needs to be made comparable or aggregated. Such a system could, for example, be used in areas other than national quality reports, such as systematic reviews and meta-analyses in the field of rehabilitation (34), or country comparisons of rehabilitation outcomes with different versions of the BI or the FIM.

The common metric established in this study has 4 major advantages over other transformation systems, such as, for example, an expert-based transformation system developed in the original ANQ ADL Score study (22) or scale equating without a reference metric, such as equating using the Leunbach model (35). First, the common metric approach is based on the ICF, the global standard for conceptualizing and describing functioning information, serving as neutral and conceptual reference to compare the content of the different tools included in the common metric (4, 36). Secondly, Rasch analysis allows an interval-based scoring and transformation table to be derived to support the calculation of valid change scores of functioning, e.g. between admission and discharge, which can inform clinical practice and research about functioning change in a quantifiable way (37). Thirdly, the common metric allows us to take the operational ranges of the 2 assessment tools into account, and can therefore operate at the level of the more detailed tool (17). Fourthly, through the ICF-based common metric, other assessment tools measuring functioning could be added in future (5). Given that this study was conducted in the context of Swiss national quality reports, only 2 assessments were included. In principle, any number of instruments can be integrated into a common metric as long as they are conceptually equivalent. Examples with more than 2 scales exist and have been published (5, 13, 20).

The comparison of the operational ranges of the 2 assessment tools showed that FIM™ covers a wider operational range than EBI. This finding can inform the choice of assessment tools and the interpretation of change scores. The basis of adding further assessment tools to the common metric would again be the linking of the respective assessment tool with the ICF and a person-equating design, in which the new tool to be added to the metric is assessed in parallel with either EBI or FIM™. As FIM™ is the assessment tool with the wider operational range, and is more widely used in rehabilitation in general, and was also previously used in other scale-equating projects, e.g. with the Barthel Index or the minimal dataset (11, 17, 38), it appears to be the choice as a linking scale.

Attention should also be paid to the equating design, as indirect transformation between 2 tools, i.e. the equating of instruments via one instrument that is already included in a transformation set, have been shown to be less precise than the direct transformation, i.e. the equating of different instruments in the same study via a common person design (35). However, indirect transformations are likely to be more feasible in terms of data collection. This reflects typical challenges in the practice of quality improvement work (39). A limitation of the study is that the dataset only covers data from 2 assessment tools and includes only neurological rehabilitation patients. The common metric would also be needed for other assessment tools or rehabilitation groups, such as that of musculoskeletal patients in the example of the Swiss quality reports. As previous studies of each assessment tool showed no substantial DIF between neurological and musculoskeletal rehabilitation patients (27, 28), the transformation table could preliminarily be applied to the musculoskeletal ANQ reports. Nevertheless, this should be validated in a future equating study including musculoskeletal data. Another limitation of the study was the restrictions of the RUMM2030 software not being able to cover total score or testlets with more than 100 thresholds (e.g. a score of 0–100), leading to some adaptions in the analyses. Likewise, the Rasch analyses had to be 2-tiered in order to enable judgement about the model fit. Furthermore, the current system does not facilitate the separation of cognitive from motor performance, as it is based on the total scores of the ADL tools, which reflect the level of reporting of the Swiss quality reports in rehabilitation. However, previous studies on EBI (27) and FIM™ (28) showed that the total scores, including motor and cognitive items, can be reported as a unidimensional construct, i.e. functional independence. Another methodological challenge was that the weight of ICF linking was not clearly defined, i.e. there is no clear cut-off when conceptual equivalency is achieved. However, with the methodological approach chosen, the conclusion drawn, based at the conceptual level on the results of the ICF linking, was supported at an empirical level by the Rasch analysis.

A strength of the present study, in addition to the above-mentioned advantages of the common metric, includes the use of the Rasch model as a basis to create the common metric. In the Rasch model the raw score is a sufficient statistic, such that there is a one-to-one correspondence between the raw score and the latent estimate, which is not the case in other item response theory or classical test theory models (16). Furthermore, the resulting estimates of the Rasch model, which build the basis for the common metric, are distribution free, given the person and item parameter separation. The common metric is therefore not sample dependent and can be applied to any relevant sample (16).

Application in practice

The ICF-based and interval-scaled common metric including the FIM™ and the EBI, as provided in this study, can be applied in national quality reports for neurological rehabilitation and, based on previous evidence with regards to DIF, for musculoskeletal rehabilitation. Application of the transformation table enables the comparison of clinics that are using different assessment tools and, at the same time, the calculation of valid change scores for patients between admission and discharge, as the transformation is based on the interval-scale level. Extension of the common metric with other functioning-related measures is possible. Possible extensions could be informed on the basis of the conducted ICF linking in comparison with the ICF Rehabilitation Set, also referred to as the ICF Generic-30, which was designed to define ICF categories as a minimal standard for reporting and assessing functioning and disability in clinical populations along the continuum of care (40).

Conclusion

The ICF-based and interval-scaled common metric provided in this study supports the assessment of patient outcomes using different ADL assessment tools (FIM™ and EBI) in clinical practice, while at the same time being able to compare the related outcomes of different clinics. It lays the groundwork for a standardized reporting system of functioning information for use in national quality improvement reports.

ACKNOWLEDGEMENTS

The authors thank Melissa Selb from the ICF Research Branch for her support in the ICF linking, Martin Brünger, Stefanie Köhn, Anna Schlumbohm and Karla Spyra from the Institute of Medical Sociology and Rehabilitation Science of the Charité – Universitätsmedizin Berlin, Germany for providing and supporting us with the data from the ANQ ADL Score Study, Luise Menzi, Head of Rehabilitation ANQ, Klaus Schmitt, former Corporate Development Director and Anke Scheel-Sailer, Head of Clinical Quality Management Research Department from the Swiss Paraplegic Centre, for their provision of valuable advice for the project. This project is part of the cumulative Dissertation of Roxanne Maritz, which is funded by the Swiss National Science Foundation's National Research Programme "Smarter Health Care" (NRP 74) within the NRP74 StARS project.

The authors have no conflicts of interest to declare.

REFERENCES

1. Stucki G, Bickenbach J. Functioning: the third health indicator in the health system and the key indicator for rehabilitation. Eur J Phys Rehabil Med 2017; 53: 134–138.
2. Stucki G, Bickenbach J, Melvin J. Strengthening rehabilitation in health systems worldwide by integrating information on functioning in national health information systems. Am J Phys Med Rehabil 2017; 96: 677–681.
3. World Health Organization (WHO). Rehabilitation in health systems: guide for action. 2019 [cited 2020 Mar 5). Available from: https://www.who.int/rehabilitation/rehabilitation-guide-for-action/en/.
4. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO, 2001.
5. Prodinger B, Tennant A, Stucki G. Standardized reporting of functioning information on ICF-based common metrics. Eur J Phys Rehabil Med 2018; 54: 110–117.
6. Haigh R, Tennant A, Biering-Sorensen F, Grimby G, Marincek C, Phillips S, et al. The use of outcome measures in physical medicine and rehabilitation within Europe. J Rehabil Med 2001; 33: 273–278.
7. Grill E, Stucki G, Scheuringer M, Melvin J. Validation of International Classification of Functioning, Disability, and Health (ICF) Core Sets for early postacute rehabilitation facilities: comparisons with three other functional measures. Am J Phys Med Rehabil 2006; 85: 640–649.
8. Prosiegel M, Böttger S, Schenk T, König N, Marolf M, Vaney C, et al. Der erweiterte Barthel - Index (EBI) – eine neue Skala zur Erfassung von Fähigkeitsstörungen bei neurologishen Patienten. Neurol Rehabil 1996; 1: 7–13.
9. Ohura T, Hase K, Nakajima Y, Nakayama T. Validity and reliability of a performance evaluation tool based on the modified Barthel Index for stroke patients. BMC Med Res Methodol 2017; 17: 131.
10. Nayar M, Vanderstay R, Siegert RJ, Turner-Stokes L. The UK Functional Assessment Measure (UK FIM+FAM): psychometric evaluation in patients undergoing specialist rehabilitation following a stroke from the National UK Clinical Dataset. PloS One 2016; 11: e0147288.
11. Li CY, Romero S, Bonilha HS, Simpson KN, Simpson AN, Hong I, et al. Linking existing instruments to develop an activity of daily living item bank. Eval Health Prof 2018; 41: 25–43.
12. Smith RM, Taylor PA. Equating rehabilitation outcome scales: developing common metrics. J Appl Meas 2004; 5: 229–242.
13. Prodinger B, Stamm T, Peterson D, Stucki G, Tennant A. Toward a standardized reporting of outcomes in hand osteoarthritis: developing a common metric of outcome measures commonly used to assess functioning. Arthritis Care Res 2016; 68: 1115–1127.
14. Victorson D., Schalet BD, Kundu S, Helfand BT, Novakovic K, Penedo F, 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: 3249–3258.
15. Oude Voshaar MAH, Vonkeman HE, Courvoisier D, Finckh A, Gossec L, Leung YY, et al. Towards standardized patient reported physical function outcome reporting: linking ten commonly used questionnaires to a common metric. Qual Life Res 2019; 28: 187–197.
16. Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011; 11: 571–585.
17. Prodinger B, O'Connor RJ, Stucki G, Tennant A. Establishing score equivalence of the Functional Independence Measure motor scale and the Barthel Index, utilising the International Classification of Functioning, Disability and Health and Rasch measurement theory. J Rehabil Med 2017; 49: 416–422.
18. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J Rehabil Med 2011; 43: 884–891.
19. Berlucchi G. Brain plasticity and cognitive neurorehabilitation. Neuropsychol Rehabil 2011; 21: 560–578.
20. Prodinger B, Tennant A, Stucki G, Cieza A, Ustun TB. Harmonizing routinely collected health information for strengthening quality management in health systems: requirements and practice. J Health Serv Res Policy 2016; 21: 223–228.
21. ANQ Nationaler Verein für Qualitätsentwicklung in Spitälern und Kliniken. Review information: rehabilitation. 2020 [cited 2020 Mar 6]. Available from: https://www.anq.ch/en/departments/rehabilitation/.
22. ANQ Nationaler Verein für Qualitätsentwicklung in Spitälern und Kliniken. Abschlussbericht Entwicklung und Validierung eines ADL–Überführungsalgorithmus auf Basis von FIM® und EBI 2018. [cited 2020 Mar 6]. Available from: https://www.anq.ch/wp-content/uploads/2018/03/ANQ_Reha_Abschlussbericht_ADL_Algorithmen.pdf.
23. Andrich D. The Polytomous Rasch Model and the equating of two instruments. In: Christensen KB, Kreiner, S, Mesbah M, editors. Rasch models in health. Hoboken: ISTE Ltd and John Wiley and Sons Inc.; 2013, p. 163–196.
24. Kolen MJ, Brennan RL. Introduction and concepts. In: Kolen MJ, Brennan RL, editors. Test equating, scaling, and linking. New York: Springer; 2004, p. 1–28.
25. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The Functional Independence Measure: a new tool for rehabilitation. Adv Clin Rehabil 1987; 1: 6–18.
26. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965; 14: 61–65.
27. Maritz R, Tennant A, Fellinghauer CS, Stucki G, Prodinger B. The Extended Barthel Index (EBI) can be reported as a unidimensional interval-scaled metric – a psychometric study. Phys Med Rehab Kuror 2019; 29: 224–232.
28. Maritz R, Tennant A, Fellinghauer C, Stucki G, Prodinger B. The Functional Independence Measure 18-item version can be reported as a unidimensional interval-scaled metric: internal construct validity revisited. J Rehabil Med 2019; 51: 193–200.
29. Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF linking rules to strengthen their potential for establishing comparability of health information. Disabil Rehabil 2019; 41: 574–583.
30. Ehlan AH, Kucukdeveci AA, Tennant A. The Rasch measurement model. In: Franco Franchignoni (editor). Research issues in physical & rehabilitation medicine. Pavia: Maugeri Foundation; 2010, p. 89–102.
31. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007; 57: 1358–1362.
32. Andrich D, Hagquist C. Real and Artificial differential item functioning in polytomous items. Educ Psychol Meas 2015; 75: 185–207.
33. Rouquette A, Hardouin JB, Vanhaesebrouck A, Sébille V, Coste J. Differential Item Functioning (DIF) in composite health measurement scale: recommendations for characterizing DIF with meaningful consequences within the Rasch model framework. PloS One 2019; 14: e0215073.
34. Stucki G, Pollock A, Engkasan JP, Selb M. How to use the International Classification of Functioning, Disability and Health as a reference system for comparative evaluation and standardized reporting of rehabilitation interventions. Eur J Phys Rehabil Med 2019; 55: 384–394.
35. Adroher ND, Kreiner S, Young C, Mills R, Tennant A. Test equating sleep scales: applying the Leunbach's model. BMC Med Res Methodol 2019; 19: 141.
36. World Health Organization, WHO. How to use the ICF – a practical manual for using the International Classification of Functioning, Disability and Health (ICF). Exposure draft for comment. 2013 (cited 2020 Mar 6). Available from: http://www.who.int/classifications/drafticfpracticalmanual2.pdf.
37. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med 2012; 44: 97–98.
38. 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. J Rehabil Res Dev 2007; 44: 467–478.
39. Needham DM, Sinopoli DJ, Dinglas VD, Berenholtz SM, Korupolu R, Watson SR, et al. Improving data quality control in quality improvement projects. Int J Qual Health Care 2009; 21: 145–150.
40. Prodinger B, Cieza A, Oberhauser C, Bickenbach J, Ustun TB, Chatterji S, et al. Toward the International Classification of Functioning, Disability and Health (ICF) Rehabilitation Set: a minimal generic set of domains for rehabilitation as a health strategy. Arch Phys Med Rehabil 2016; 97: 875–884.