



RASCH ANALYSIS OF THE UK FUNCTIONAL ASSESSMENT MEASURE IN A SAMPLE OF PATIENTS WITH TRAUMATIC BRAIN INJURY FROM THE UK NATIONAL CLINICAL DATABASE

Lynne TURNER-STOKES, DM, FRCP^{1,2}, Oleg N. MEDVEDEV, PhD^{3,4} and Richard J. SIEGERT, PhD⁴

From the ¹King's College London, Department of Palliative Care, Policy and Rehabilitation, Faculty of Life Sciences and Medicine, ²Regional Hyper-acute Rehabilitation Unit, Northwick Park Hospital, London North West Healthcare NHS Trust, London, UK, ³University of Waikato, School of Psychology, and ⁴Auckland University of Technology, School of Public Health & Psychosocial Studies and School of Clinical Sciences, Auckland, New Zealand

Objective: To determine whether the UK Functional Assessment Measure (UK FIM+FAM) fits the Rasch model in patients with complex disability following traumatic brain injury.

Design: Psychometric evaluation including preliminary exploratory and confirmatory factor analyses followed by Rasch analysis.

Participants: A multicentre UK national cohort of 1,956 patients admitted for specialist rehabilitation following traumatic brain injury.

Results: The suitability of the Partial Credit Model was confirmed by the likelihood-ratio test (χ^2 (df86)=7,325.0, $p<0.001$). Exploratory and confirmatory factor analyses supported 3 factors (Motor, Communication, Psychosocial). Rasch analysis of the full scale incorporating the 3 factors as super-items resulted in an acceptable overall model fit (χ^2 (df24)=36.72, $p=0.05$) and strict uni-dimensionality when tested on a sub-sample of $n=320$. These results were replicated in a full sample ($n=1,956$) showing uni-dimensionality and good reliability with Person Separation Index=0.81, but item trait interaction was significant due to the large sample size. No significant differential item functioning was observed for any personal factors. Neither uniform rescoring of items nor exclusion of participants with extreme scores improved the model fit.

Conclusion: The UK FIM+FAM scale satisfies the Rasch model reasonably in traumatic brain injury. A conversion table was produced, but its usefulness in clinical practice requires further exploration and clinical translation.

Key words: Rasch analysis; functional assessment measure; rehabilitation; outcome; traumatic brain injury.

Accepted Jun 14, 2019; Epub ahead of print Jun 27, 2019

J Rehabil Med 2019; 51: 566–574

Correspondence address: Lynne Turner-Stokes, King's College London, Department of Palliative Care, Policy and Rehabilitation, Faculty of Life Sciences and Medicine, London, UK. E-mail: lynne.turner-stokes@doctors.org.uk

Outcome measurement is a key requirement for rehabilitation services to support comparison of programmes, practices and populations. Traumatic brain injury (TBI) poses particular challenges

LAY ABSTRACT

The UK Functional Assessment Measure (UK FIM+FAM) is a scale used to evaluate the outcome of rehabilitation after acquired brain injury as part of the UK Rehabilitation Outcomes Collaborative (UKROC). The FIM+FAM is administered by clinicians and measures how well people are recovering their independence. It has 30 separate questions concerning a person's physical (e.g. eating, toileting, bathing) and cognitive (e.g. expression, memory, safety awareness) independence. Data from 1,956 UK patients who had survived a traumatic brain injury were analysed. The main purpose of this study was to establish that the FIM+FAM meets some of the stringent technical standards for clinical and scientific measurement. A statistical method known as Rasch analysis was used to demonstrate that the FIM+FAM produces accurate and reliable scores. The results showed that the FIM+FAM is a practical and useful scale for measuring change in people in rehabilitation after a traumatic brain injury.

for outcome evaluation in rehabilitation. As well as improvement in independence for activities of daily living, measurement tools need to encompass changes in cognitive, communicative and psychosocial function, which are often the main factors that limit independence following TBI.

The Functional Independence Measure (FIMTM) (1, 2) is an 18-item measure of functional independence, which is widely used as a measure of outcome from rehabilitation. Although 5 of the items address elements of communication and cognitive function, its predominant focus is on physical function and it has well recognized "ceiling" effects (3). The Functional Assessment Measure (FAM) was developed in the 1990s specifically for more ambulant patients with TBI. It adds 12 items to the FIM scale, mainly addressing cognitive and psychosocial function, with the intent of extending the range of difficulty in this group (3, 4). From a scaling perspective, the FAM items as a whole may have little impact on the ceiling effects of the FIMTM in higher functioning patients (5), but from a clinical perspective they may still provide useful information about the less visible aspects of neurological disability for patients at all levels (6).

METHODS

Adapted from the original US version, the UK Functional Assessment Measure (UK FIM+FAM) was first published in 1999 (7). The 30-item scale comprises 16 motor items (including self-care and mobility) and 14 cognitive items (including communication, cognition and psychosocial adjustment).

The UK Rehabilitation Outcomes Collaborative (UKROC) (8) provides the national clinical database for all specialist in-patient rehabilitation services in England. These services cater for a selected population of patients with severe complex neurological disability. The UK FIM+FAM is the principal measure of functional outcome within the dataset and, since 2013, its collection is a mandated requirement for reimbursement within these services. It is therefore pertinent to understand its psychometric properties in this context. In addition, transformation of the scale from ordinal to interval-level data (using techniques such as Rasch analysis (9–11)) confers theoretical advantages for clinical practice, potentially increasing sensitivity by stretching the existing ogive raw score points at both the upper and lower margins, as well as enabling the calculation of valid change scores.

Traditional psychometric evaluations of the UK FIM+FAM in both a general neuro-rehabilitation cohort and in stroke patients (12, 13) have shown the UK FIM+FAM to have 3 distinct subscales: Motor, Communicative, and Psychosocial. Two studies from the 1990s have explored the linear Rasch transformation of the original US FIM+FAM scale following TBI and showed partial conformation to the Rasch model (14, 15). However, approaches to Rasch analysis have changed significantly in the last 20 years (16), and as yet there are no published Rasch analyses of the UK FIM+FAM scale in patients with TBI.

Our recent Rasch analysis of the UK FIM+FAM in patients with complex disability following stroke (17) provided evidence for uni-dimensionality. It showed that the best fit was achieved where responses for the 16-item "Motor", 5-item "Communicative" and 9-item "Psychosocial" subscales were summarized into 3 super-items, and were split for left and right hemisphere stroke location due to differential item functioning (DIF). This approach satisfied the expectations of the Rasch model without the need for re-scoring of item thresholds or exclusion of extreme scores.

The aim of this study was to determine whether the UK FIM+FAM fits the Rasch model in patients with complex disability following TBI in a similar manner to patients with stroke, and if so, to derive a conversion table of Rasch-transformed scores for potential clinical use.

Setting and participants

In the UK, rehabilitation services are broadly categorized on 3 levels (18). Following TBI, the majority of patients will have relatively simple rehabilitation needs and make a good recovery with the support of their local (Level 3) rehabilitation services. However, a smaller number have more complex disability requiring specialist rehabilitation in either local district (Level 2) or regional (Level 1) rehabilitation services. The UKROC database provides the national clinical database for all Level 1 and 2 rehabilitation services in England. The study sample therefore represents a selected group of adults with complex disability following severe brain injury.

Principles of approach

Where an instrument is already established in clinical use and the content and ordinal scoring system have some meaning for clinicians, a balance may need to be found between maintaining the integrity of the scale and finding the best fit solution for a transformed scale. In addition, patients with complex disability following TBI form a more diverse patient group than stroke patients, with deficits ranging from ambulant patients with highly challenging behaviours to severe physical disability. The diversity of this sample represents a challenge for the psychometric properties of the FIM+FAM. Hence, we expected that this analysis might not fit the Rasch model quite as readily as in more homogeneous conditions.

UK FIM+FAM

The 30 items of UK FIM+FAM are each scored on a 7-point ordinal scale as follows: 1 (Total assistance); 2 (Maximal assistance); 3 (Moderate assistance); 4 (Minimal assistance); 5 (Supervision/set-up); 6 (Independent with device); and 7 (Fully independent). A category of 6 or 7 implies no help from another person, whilst assessment for categories 1 to 4 is often based on frequency of intervention. Like the US version, the FIM components of the UK FIM+FAM are retained in order to maintain comparability for that component of the scale with units that use the FIM only. Amongst the FAM items, 3 are structurally different in the UK version (1999) compared with the original US version (1994): "Concentration" replaces "Attention"; "Safety awareness" replaces "Safety judgement"; and "Use of leisure time" replaces "Employability". The rationale for these differences is described in the 1999 source paper for the UK FIM+FAM.

Data source

The data source for this analysis was the UKROC database, which was initially set up by a National Institute for Health Research Programme Grant (19). It is now commissioned directly by NHS England to provide the national clinical and commissioning database for specialist inpatient rehabilitation in England. The dataset comprises socio-demographic and clinical data as well as information on rehabilitation needs, inputs and outcomes on admission and discharge from inpatient rehabilitation. Reporting of UK FIM+FAM data has been a commissioning requirement for Level 1 and 2 rehabilitation

services since April 2013. Specialist Level 1 and 2 rehabilitation services have a mean length of stay of approximately 90 days (SD 66) (20). The UK FIM+FAM is usually completed for each patient within 10 days of admission and during the last week before discharge, to evaluate the functional gains made during the episode of care.

Sampling

We extracted the cohort of all 1,956 TBI patients consecutively admitted to the 58 Level 1 and 2 specialist neurorehabilitation centres that submitted data to the UKROC database between 1 January 2010 and 30 May 2016, for whom a UK FIM+FAM score was available both at admission and discharge from the rehabilitation programme. A similar process for data extraction and analysis was used to that in the previous publications for stroke (12, 17), as summarized in Fig. 1.

The scores are expected to be lower (more dependent) on admission and higher (more independent) on discharge. Therefore, to ensure the full range of the response category necessary to evaluate reliability, and in line with previous analyses (17), we pooled admission and discharge scores from the complete sample of $n=1,956$ into a single dataset. In order not to violate the Rasch assumption of local independence between observations, we included only one time-point for each patient, with both time-points equally represented. For the main dataset the random sampling function in SPSS was used to divide the dataset into approximately equal halves, giving 960 admission scores and 996 discharge scores. A mirror dataset was prepared using the remaining scores (996 admission and 960 discharge scores) for confirmatory factor analysis (CFA) of an independent sample.

Overview of analysis

The key steps were as follows:

- An initial exploratory factor analysis (EFA) was conducted on the main dataset ($n=1,956$) using the IBM Statistical Package for Social Sciences (SPSS) v23 software. A principal components analysis (PCA) was applied with varimax rotation. The decision regarding the number of factors to rotate was based on consideration of the number of factors with Eigenvalues >1.0 and visual inspection of the scree plot.
- CFA was conducted on the mirror dataset ($n=1,956$) using the IBM Analysis of Moment Structures (AMOS) software, which uses visual statistical software for CFA. The quality of the model fit was assessed on 5 indices: χ^2/df ratio, p -value, root-mean-square error of approximation (RMSEA), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Normed Fit Index (NFI). In general, excellent fit is represented by a model where $\chi^2/df < 2.0$, $p > 0.05$, RMSEA < 0.08 , CFI ≥ 0.95 , TLI ≥ 0.95 and NFI ≥ 0.95 . We tested the best factor model identified in the EFA.
- Rasch analysis was conducted on the main dataset ($n=1,956$) using RUMM2030 software (23) incorporating the method detailed below. Sample size for Rasch analysis depends on the degree of precision required for item calibration (16). Linacre et al. recommends a minimum sample size of 108–243, depending on how well the subjects are targeted to the scale, although for “high stakes settings” (e.g. with assessments contributing to clinical diagnosis) a minimum sample is recommended of 20 cases per item or 250 participants, whichever is the greater (24). This would mean that the FIM+FAM scale would require up to 600 cases if the intended use was for individual patient assessment (16). However, larger datasets ($n > 500$) are associated with inflated χ^2 fit statistics in RUMM2030 (25, 26) and a smaller sample size was used for the χ^2 tests of a similar study in stroke (e.g. $n=320$) (17).

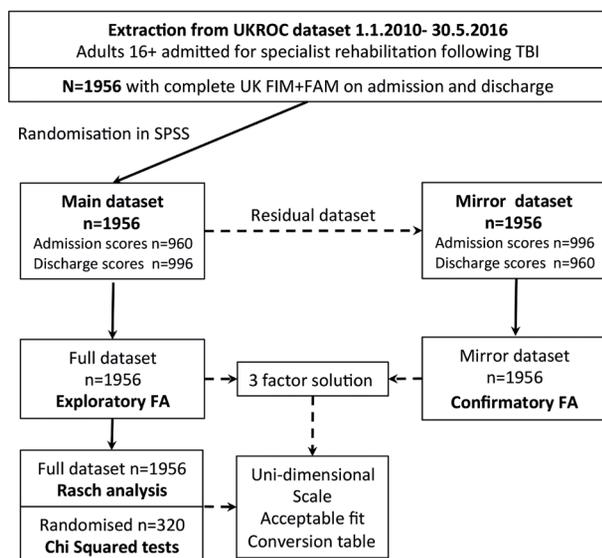


Fig. 1. Summary of data sampling. Of the 1,956 consecutively admitted traumatic brain injury (TBI) patients with complete UK Functional Assessment Measure (UK FIM+FAM) data, admission and discharge scores were randomly selected (using 1 time-point only for each patient) to form the main dataset for exploratory factor analysis (EFA) and Rasch analysis. A mirror dataset was prepared using the remaining scores for confirmatory factor analysis of an independent sample. UKROC: UK Rehabilitation Outcomes Collaborative; SPSS: IBM Statistical Package for Social Sciences; confirmatory FA: confirmatory factor analysis.

Rasch analysis

The most suitable type of Rasch model for the analysis was determined by the likelihood-ratio test examining the assumption of the Rating Scale Model acting as a null hypothesis that distribution of item thresholds across individual scale items is the same. If the likelihood-ratio test is significant it rejects the Rating Scale Model. Since we are examining the full scale and not specific groups of items, there is no alternative for the unrestricted Partial Credit Model if the Rating Scale Model is rejected. The summary statistics of the Rasch model were assessed based on mean item and person location, fit residual, item-trait interaction χ^2 test/ p -value. A scale with the items ideally targeted to the population has the mean person location and the mean item location approximately at zero logits (SD 1), and item and person distributions that mirror one another. The item-trait interaction reflects the fit of the data to the model's expectations; a significant p -value of < 0.05 indicates inadequate fit to the model. Reliability was estimated using Person Separation Index (PSI). The PSI is a measure of scale ability to discriminate between persons with different trait levels. Its values can be interpreted similar to Cronbach's alpha used in classical test theory (27); values above 0.7 are required for group use and above 0.8 for individual assessment (although values of over 0.9 are preferred).

The Rasch analysis was carried out in 2 main analytical pathways using the full main dataset, but, as the subjects were found to be well-targeted to the scale (see Results section), item trait and χ^2 tests were conducted on the reduced subset ($n=320$) to account for the effects of sample size. In the first analysis, all

Table I. The UKROC: trauma brain injury population and the Rasch random sample characteristics

	UKROC Study sample <i>n</i> = 1,956	Random sample (Rasch analysis) ^a <i>n</i> = 320
Age, <i>n</i> (%)		
16–40 years	689 (35.2)	114 (35.6)
41–65 years	830 (42.4)	140 (43.8)
> 65 years	437 (22.3)	66 (20.6)
Male, <i>n</i> (%)	1,466 (74.9)	241 (75.3)
Ethnicity, <i>n</i> (%)		
White	1,463 (74.8)	236 (73.8)
Asian/Asian British	88 (4.5)	13 (4.1)
Black/Black British	28 (1.4)	5 (1.6)
Other	138 (7.1)	23 (7.2)
Unknown	239 (12.2)	43 (13.4)
Length of stay, days, mean (SD)	78.2 (68.2)	77.5 (71.3)
Diagnosis localization, <i>n</i> (%)		
Global/diffuse	640 (32.7)	110 (34.4)
Left hemisphere	480 (24.5)	65 (20.3)
Right hemisphere	428 (21.9)	75 (23.4)
Frontal	408 (20.9)	70 (21.9)
FIM+FAM, mean (SD)		
Total FIM on admission	66.6 (35.0)	68.0 (35.5)
Total FAM on admission	107.7 (52.6)	109.5 (53.3)
Total FIM on discharge	91.9 (35.5)	93.0 (35.7)
Total FAM on discharge	149.7 (55.8)	151.6 (55.9)

^aRandom sample extracted from the dataset (*n* = 1,956) derived across admission and discharge values, so that each patient is only in the dataset once, but both time-points are equally represented. UKROC: UK Rehabilitation Outcomes Collaborative; SD: standard deviation; FIM: Functional Independence Measure; FAM: Functional Assessment Measure.

30 items were fitted to the Rasch Model without adjustment of any kind. The second used a super-item approach based on the method used by Lundgren Nilsson and colleagues (16) to solve local dependency issues between items of distinct domains. Locally dependent items were combined into 3 super-items based on the 3-factor structure identified by steps 1 and 2 above.

In line with our previous analysis in stroke (17), item calibrations and DIF were explored across potential person factors including age group (16–32, 33–49, 50–65, > 65 years), sex, diagnosis subcategory (global/diffuse, frontal, left hemisphere, right hemisphere), and time-point (admission or discharge) and patient dependency level based on The Northwick Park Dependency Score (28). The uni-dimensionality of the scale was assessed using PCA of the residuals (29). This involves comparing person estimates of the 2 groups of items with highest positive and negative loadings on the first principal component of residuals after controlling for the latent Rasch factor using a *t*-test procedure. Evidence of uni-dimensionality is obtained if the percentage of significant *t*-tests is below 5% or the lower bound of a binomial confidence interval computed for the number of significant *t*-tests overlaps the 5% cut-off point (29). A significance value of 0.05 was used throughout.

RESULTS

The socio-demographic and clinical characteristics of the Rasch study sample, and the UKROC population from which it was drawn, are shown in Table I. The

Table II. Principal component loadings after varimax rotation

UK FIM+FAM item	Single component* 60%	2-factor solution		3-factor solution		
		70%		73%		
Variance accounted for:		Motor	Cognitive	Motor	Psychosocial	Communication
Eating	0.827	0.634	(0.533)	0.703		
Swallowing	0.763	0.550	(0.529)	0.631		
Grooming	0.904	0.709	(0.566)	0.719		
Bathing	0.887	0.807		0.788		
Dressing upper body	0.897	0.789		0.802		
Dressing lower body	0.883	0.854		0.839		
Toileting	0.879	0.869		0.868		
Bladder	0.832	0.747		0.760		
Bowels	0.860	0.778		0.799		
Bed transfers	0.850	0.914		0.900		
Toilet transfers	0.847	0.915		0.903		
Bath transfers	0.809	0.894		0.859		
Car transfers	0.682	0.753		0.610		
Locomotion	0.810	0.880		0.839		
Stairs	0.661	0.779		0.635		
Community mobility	0.598	0.580		0.408		
Comprehension	0.762		0.799			0.640
Expression	0.770		0.780			0.655
Reading	0.677		0.665			0.759
Writing	0.670		0.615			0.723
Speech Intelligibility	0.732		0.606			0.572
Social interaction	0.700		0.772		0.714	
Emotional status	0.586		0.650		0.663	
Adjustment	0.681		0.814		0.832	
Leisure activities	0.746		0.699		0.669	
Problem solving	0.768		0.789		0.757	
Memory	0.710		0.812		0.813	
Orientation	0.739		0.821		0.781	
Concentration	0.762		0.805		0.784	
Safety awareness	0.704		0.752		0.779	
Cronbach's alpha	0.976	0.976	0.954	0.976	0.913	0.943

Loadings < 0.5 were suppressed. Pearson Item – total correlations for the single scale were all significant at *p* < 0.001. FIM: Functional Independence Measure; FAM: Functional Assessment Measure.

full dataset, and the Rasch sub-sample of $n = 320$ were similar on all parameters.

The results of EFA are summarized in Table II. All items loaded strongly onto a single first component (all loadings > 0.58) with Cronbach's alpha 0.976, accounting for 60% of the variance. Four factors had eigenvalues > 1.0 , but the fourth contained only one item ("Community mobility"), which is rarely assessed and so rated 1 ("not assessed") by scoring convention. The remaining 3 factors (Motor (16 items), Psychosocial (9 items) and Communication (5 items)) together accounted for 73% of the variance. As suggested by a previous factor analysis in TBI patients, a 2-factor model was also explored, which accounted for 70% of the variance.

CFA for a 3-factor model showed a marginal fit as shown in Figure 2. This model included correlated of error terms for nine pairs of items with the largest modification indices. The fit indices did not meet the criteria for an excellent fit, but were all in the borderline acceptable range (RMSEA 0.094, CFI=0.926,

TLI=0.918, NFI=0.922) While the χ^2/df ratio was large (18.103), giving a highly significant p -value (0.000), this is more likely to be a reflection of the large sample size than the model fit (30). The fit was only marginally improved by further correlation of error terms and thus we proceeded to Rasch analysis.

Rasch analysis

The suitability of the Partial Credit Model for Rasch analysis was confirmed by the significant likelihood-ratio test ($\chi^2 (df86) = 7,325.0, p < 0.001$) Table III includes fit statistics for individual items together with response frequencies for each category within the main dataset ($n = 1,956$) and for the domain scores presented for both the full dataset and the sub-sample ($n = 320$). A preponderance of scores at the ends of the range is expected for a dataset that includes both admission and discharge data, and indeed there was a greater preponderance of response category scores at the lowest end of the scale for the admission sample, and upper end in the discharge sample. Table IV summarizes the overall fit statistics from the main stages of the Rasch analysis.

First analytical pathway (all 30 items). The initial analysis of the full 30-item scale is marked by satisfactory reliability (PSI=0.94). However, the overall model fit was poor with significant item-trait interaction. At the individual item level, 14 out of 30 items showed significant misfit to the Rasch model on the $n = 320$ sub-sample (Table III).

Second analytical pathway (3 super-items). The second analytical pathway was conducted using 3 super-items created by combining the items within the Motor, Communication and Psychosocial domains identified from the factor analyses. Prior to combining domain items into super-items residual correlation matrix was examined. Residual correlations exceeding the cut-off point of 0.20 above the mean of all residual correlations were found for 59 pairs of items in Motor, 5 pairs in Communication, and 26 pairs in Psychosocial domain. Rasch analysis of 3 super-items resulted in acceptable overall model fit ($\chi^2 (df24) = 36.72, p = 0.05$), strict uni-dimensionality and no local dependency when tested on the sub-sample ($n = 320$) (Table IV, Pathway 2). The proportion of common error-free variance $A = 0.88$ was marginally below 0.90 suggesting that use of transformation table based on this 3 super-item solution may be preferable to the ordinal scale for calculation of change scores. This analysis was replicated with the full dataset ($n = 1,956$) showing good reliability PSI of 0.81,

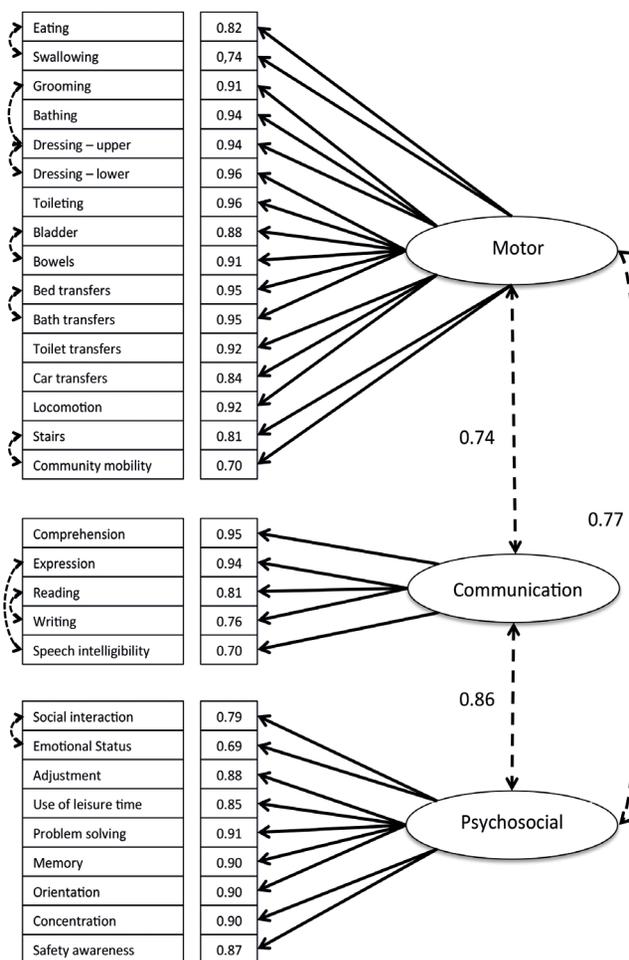


Fig. 2. Summary of findings from confirmatory factor analysis. CFA confirmed a 3-factor solution. Correlations between the factors ranged from 0.74 to 0.86.

Table III. Frequency distribution of responses and Rasch model fit statistics for the UKFIM+FAM items (*n* = 320). See Table S1¹ for an equivalent table for individual items the full dataset (*n* = 1,956)

Description	Location	Fit Residual	χ^2 **	Frequency distribution across scoring categories							% Cat 1+7
				Cat 1	Cat 2	Cat 3	Cat 4	Cat 5	Cat 6	Cat 7	
Motor											
Eating	-0.56	-0.71	8.34	37	7	7	8	30	35	170	70.4
Swallowing*	-0.69	4.23	29.30	28	9	7	3	17	27	203	78.6
Grooming*	-0.21	-4.56	34.33	43	16	16	23	44	47	105	50.3
Bathing*	0.05	-4.54	33.87	50	24	14	36	47	45	78	43.5
Dressing – upper*	-0.23	-4.74	33.27	43	21	13	24	39	36	118	54.8
Dressing – lower*	-0.05	-4.58	33.17	51	22	20	26	40	33	102	52.0
Toileting*	-0.15	-5.24	31.61	58	17	17	10	25	32	135	65.6
Bladder	-0.25	-1.71	7.08	57	12	12	11	16	27	159	73.5
Bowel*	-0.24	-2.80	11.17	58	10	14	9	15	26	162	74.8
Bed transfers*	-0.17	-3.62	15.82	55	9	23	13	30	36	128	62.2
Toilet transfers*	-0.12	-3.59	18.48	60	8	20	15	28	37	126	63.3
Bath transfers	0.10	-0.49	3.12	72	10	16	20	42	38	96	57.1
Car transfers	0.29	0.55	8.11	112	7	11	21	13	31	99	71.8
Locomotion	0.08	-1.53	6.78	72	12	6	12	40	57	94	56.7
Stairs	0.54	0.79	8.67	138	4	4	4	24	60	60	67.3
Community mobility	0.93	1.25	1.99	132	28	39	10	29	27	29	54.8
Communication											
Comprehension	-0.42	-0.41	2.06	22	16	16	14	76	68	82	35.4
Expression	-0.36	-0.30	5.52	29	16	15	18	55	66	95	42.2
Reading*	0.02	4.03	38.31	59	8	10	23	60	45	89	50.3
Writing*	0.23	2.90	33.34	84	10	22	31	33	34	80	55.8
Speech intelligibility	-0.67	2.09	7.68	20	11	18	11	35	40	159	60.9
Psychosocial											
Social interaction*	-0.36	3.55	18.71	24	17	16	22	52	62	101	42.5
Emotional status*	0.01	9.35	89.26	45	33	18	16	37	71	74	40.5
Adjustment to limitations	0.39	1.14	7.76	51	50	40	31	40	45	37	29.9
Use of leisure time	0.38	0.52	1.61	56	44	29	21	39	69	36	31.3
Problem solving	0.59	-0.68	5.16	73	32	27	32	53	51	26	33.7
Memory*	0.34	2.67	16.23	58	48	29	32	36	47	44	34.7
Orientation	-0.03	0.64	4.77	47	36	21	29	24	41	96	48.6
Concentration	0.04	0.42	1.69	33	42	27	22	59	51	60	31.6
Safety awareness	0.54	-0.42	4.17	54	77	34	18	32	42	37	31.0
Domain subtests											
Motor	0.01	-1.28	11.43								
Communication	-0.07	-0.08	18.95								
Psychosocial	0.06	0.99	6.34								

*Significant misfit to the Rasch model (*p* < 0.05, Bonferroni-adjusted); Response options in bold were endorsed by less than 10 people.

** χ^2 degrees of freedom for items (df=4) and for subtests (df=8)

Percent cat 1+7: response category scores in the lower and upper end of the scale

(Cronbach’s alpha=0.78), strict uni-dimensionality (<5% *t*-tests significant at <0.05), and no local dependency. However, the item-trait interaction was significant (χ^2 (df24)=159.1, *p*<0.001) confirming the expected inflation of χ^2 statistics due to large sample size. As with our previous analysis for stroke (17), neither uniform re-scoring of items nor exclusion of participants with extreme scores improved the fit of the model, so these data are not presented.

Analysis of DIF showed no significant DIF by any of the personal factors. At the individual super-item level there were no signs of local dependency and all showed acceptable fit residual and χ^2 values (Table III). Fig. S1¹ shows the person-item threshold distribution for the final analysis with the full sample and indicates

¹<http://www.medicaljournals.se/jrm/content/?doi=10.2340/16501977-2580>

Table IV. The UK FIM FAM: Rasch model summary statistics (overall fit of the scale)

UK FIM+FAM	Item – trait interaction			Uni-dimensionality (Sig. <i>t</i> -tests in %)	Local dependency
	χ^2 /df	<i>p</i> -value	PSI		
Pathway 1: Initial all 30 items					
Sub-sample (<i>n</i> = 320)	521.36/120	< 0.001	0.94	No (> 35)	
Full dataset (<i>n</i> = 1,956)	2,253.69/120	< 0.001	0.94	No (> 35)	Yes
Pathway 2: Three super-items					
Sub-sample (<i>n</i> = 320)	36.72/24	0.05	0.78	Yes (2.19)	
Full dataset (<i>n</i> = 1,956)	159.11/24	< 0.001	0.81	Yes (2.51)	No

UK FIM+FAM; UK Functional Assessment Measure; df: degrees of freedom; PSI: Person Separation Index (without extremes). ^aSuper-item 1=Communication; Super-item 3 is Psychosocial domain. FIM: Functional Independence Measure; FAM: Functional Assessment Measure.

Table V. The UK FIM+FAM conversion table based on raw and interval scores

Raw Score	Interval Scale										
30	30.0	60	113.0	90	119.1	120	121.9	150	124.6	180	130.0
31	30.0	61	113.4	91	119.2	121	121.9	151	124.6	181	130.4
32	60.0	62	113.7	92	119.3	122	122.2	152	124.8	182	130.7
33	74.8	63	114.0	93	119.4	123	122.1	153	124.9	183	131.1
34	82.1	64	114.3	94	119.6	124	122.3	154	124.9	184	131.5
35	89.2	65	114.5	95	119.7	125	122.4	155	125.0	185	131.8
36	90.8	66	114.9	96	119.8	126	122.3	156	125.2	186	132.3
37	93.5	67	115.1	97	119.9	127	122.5	157	125.3	187	132.8
38	95.8	68	115.3	98	120.0	128	122.6	158	125.5	188	133.2
39	97.6	69	115.6	99	120.1	129	122.7	159	125.6	189	133.8
40	99.2	70	115.8	100	120.2	130	122.8	160	125.7	190	134.4
41	100.6	71	116.0	101	120.3	131	122.9	161	125.9	191	135.0
42	101.9	72	116.2	102	120.5	132	122.9	162	126.0	192	135.6
43	102.9	73	116.4	103	120.4	133	122.9	163	126.2	193	136.3
44	103.9	74	116.6	104	120.6	134	123.0	164	126.3	194	137.0
45	104.8	75	116.8	105	120.7	135	123.2	165	126.5	195	137.9
46	105.6	76	117.0	106	120.8	136	123.1	166	126.6	196	138.7
47	106.3	77	117.2	107	120.9	137	123.3	167	126.9	197	139.8
48	107.0	78	117.4	108	121.0	138	123.3	168	127.0	198	140.8
49	107.7	79	117.5	109	120.9	139	123.4	169	127.2	199	142.0
50	108.3	80	117.7	110	121.1	140	123.5	170	127.4	200	143.3
51	108.8	81	117.8	111	121.2	141	123.6	171	127.6	201	144.9
52	109.4	82	118.0	112	121.3	142	123.7	172	127.9	202	146.6
53	109.9	83	118.2	113	121.3	143	123.8	173	128.0	203	148.7
54	110.3	84	118.3	114	121.5	144	123.9	174	128.3	204	150.4
55	110.8	85	118.5	115	121.6	145	124.0	175	128.5	205	154.0
56	111.2	86	118.6	116	121.5	146	124.1	176	128.8	206	157.8
57	111.6	87	118.8	117	121.6	147	124.2	177	129.1	207	162.9
58	112.0	88	118.9	118	121.8	148	124.3	178	129.4	208	170.5
59	112.4	89	119.0	119	121.8	149	124.4	179	129.7	209	184.5
										210	210.0

excellent targeting of the sample by the items thresholds with the person mean of -0.04 (SD 0.52). This justifies using a smaller sub-sample ($n=320$) to estimate a χ^2 statistic sensitive to the sample size. Table V shows the conversion table for the full-scale model. We then correlated the interval scores from Table V with the interval scores obtained for the same FIM+FAM raw score in

our previous Rasch analysis with a UKROC stroke dataset to determine whether a single conversion table could suffice. Fig. 3 shows a scatterplot illustrating the relationship between ordinal summed raw scores and Rasch interval scores comparing TBI with left and right strokes. To investigate whether separate conversion tables for TBI and stroke patients were necessary, we pooled our previously reported stroke samples ($n=680$ left; $n=638$ right stroke) (17) with a randomly selected TBI sample of comparable size ($n=680$) and conducted DIF analysis for 3 super-item solution. Significant DIF was identified between left stroke and TBI for all 3 super-items: Motor ($F(1, 1316)=152.03, p<0.001$), Communication ($F(1, 1316)=59.84, p<0.001$) and Psychosocial ($F(1, 1316)=137.94, p<0.001$). Significant uniform DIF was also identified between right stroke and TBI samples for the Communication ($F(1, 1358)=8.64, p=0.003$) and Psychosocial ($F(1, 1358)=19.40, p<0.001$) super-items. DIF plots (A-F) are included in Fig. S2A-F. These would suggest that separate conversion tables for stroke and TBI are necessary. However, as demonstrated by Fig. 3, the differences lie predominantly in the middle range of the scale and further work is required to determine whether they are sufficiently meaningful at a clinical level to justify separate transformation tables for the different conditions.

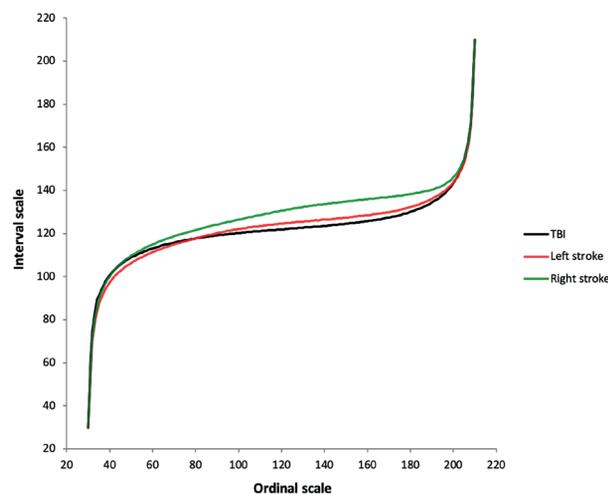


Fig. 3. Scatterplot of interval vs ordinal data. A scatterplot illustrating the relationship between the UK Functional Assessment Measure (UK FIM+FAM) ordinal summed raw scores and Rasch interval scores comparing traumatic brain injury (TBI) with left and right strokes.

DISCUSSION

This study represents the first Rasch analysis of the UK FIM+FAM scale in patients with TBI. The aim was to determine whether the instrument fits the Rasch model in a similar manner to patients with stroke, and if so, to derive a conversion table of Rasch-transformed scores for potential clinical use.

The best fit model was obtained when items were combined into the 3 super-items, and the scale fitted the Rasch model reasonably without any item deletion, amendment or rescoring. However, as with the stroke patients, at both ends of the scale the relationship between the Rasch measure and the raw scores is extremely steep, while in the middle of the scale the distribution of the scores is notably "flat". Table V also confirms that the difference in the summed raw scores between 97 and 197 (100 units on the horizontal axis in Fig. 3) corresponds to a difference in the summed Rasch interval admission scores between 120 and 140 (i.e. just 20 units on the vertical axis). That many raw score points are to be found within a narrow logit range of functioning reflects the fact that the thresholds from different items fall closely together across this functional space and presents a good example of the tension between a Clinimetric and Psychometric perspective (31). The clinical information is important at individual level, as patients are improving on tasks within a clinical framework that is recognisable by clinicians, but the psychometric evidence is that such improvement (i.e. each additional raw score point) is marginal in the overall scale of functioning, and could give a misleading impression of the speed of recovery. The opposite is true for those at the margins of the scale. Linearisation thus confers the potential advantage of greater discrimination in the upper and lower ends, together with more robust statistical properties – for example enabling calculation of valid change scores, which may be important for research, but transformed scores are less recognisable by clinicians, which may explain their limited uptake to date in clinical settings. We suggest that, while the Rasch-transformed scores may be useful for research purposes, their widespread adoption into routine clinical practice will require further work at the translational level.

Because people were expected to have a preponderance of lower scores on admission and higher scores on discharge, our main Rasch analysis was conducted on a mixed admission and discharge sample, to ensure that the full range of the response categories were represented. As illustrated in Table III, this resulted in a preponderance of scores in the lowest and highest response categories in the sample, which effectively creates a U-shaped distribution, as opposed to the more normal distribution that would be expected in a

cross-section of a disabled population. This bi-modal distribution reflects the chosen time-points at each end of the programme, rather than being a reflection of the scaling properties of the UK FIM+FAM (i.e. people not being able to endorse answers in the middle range of the scale). However, while the scarcity of mid-range scores affected the PSI by inflating measurement error, it should not have any influence on the Rasch logistic model shape because the Rasch model has no distributional assumptions. Given the extreme (bi-modal) distribution of the TBI scores, the PSI value of 0.81 found in this study may be considered satisfactory for group analyses, but use for individual assessment warrants further clinical testing.

This relationship between study sample distribution and the Rasch model fit is an interesting finding that deserves further exploration with psychometric techniques, including generalizability theory (32). The floor and ceiling effects of the UK FIM+FAM are well recognized for the population as a whole (14, 15). However, as with many scales used for outcome evaluation in rehabilitation, the level descriptors were designed to capture the types of change that may be expected to occur within an inpatient rehabilitation setting, as the patient progresses to the level of independence required for transition to the community. In this context the range is quite suitable for the population of inpatients with complex disabilities requiring specialist rehabilitation following acquired brain injury (which is the setting in which it is used in the UK(20)). But, because the FIM+FAM is rated only on admission and discharge in UK clinical settings, there is limited opportunity to capture progression through the range of scores that may be expected to occur during the rehabilitation journey. Were the instrument to be rated at frequent intervals throughout the programme (as occurs with the FIM in some US rehabilitation programmes) one might expect a more normal distribution of scores. In view of the time commitment, this is not feasible in a busy clinical setting. A more normal-shaped distribution could be also achieved by purposive selection, but this would not reflect normal clinical experience.

Limitations

In addition to the limitations above, the authors recognize that the UKROC population is a selected group of patients with severe TBI and highly complex rehabilitation needs requiring specialist rehabilitation. The findings may not therefore not be generalizable to the overall TBI population. However, they would be relevant to other groups of TBI patients requiring treatment in specialist or tertiary centres, which are probably the main services that are likely to be using the UK FIM+FAM as an outcome measure.

Conclusion

Our results suggest that UK FIM+FAM meets the Rasch model expectations, confirming that it has acceptable properties as an interval scale. Further work is required to determine whether the use of transformed scores confers significant benefit in a clinical setting.

ACKNOWLEDGEMENTS

The authors would like to express their thanks to all of the patients and clinicians who contributed to the UKROC dataset. Many thanks to Heather Williams and Keith Sephton for help with data extraction and cleaning.

This manuscript presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (RP-PG-0407-10185). The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

The authors have no conflicts of interest to declare.

REFERENCES

1. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The Functional Independence Measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987; 1: 6–18.
2. Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Tashman JS. A uniform national data system for medical rehabilitation. In: Fuhrer JM, editor. *Rehabilitation outcomes: analysis and measurement*. Baltimore: Brookes; 1987, p. 137–147.
3. Hall KM, Mann N, High WMJ, Wright J, Kreutzer JS, Wood D. Functional measures after traumatic brain injury: ceiling effects of FIM, FIM+FAM, DRS, and CIQ. *J Head Traum Rehabil* 1996; 11: 27–39.
4. Hall KM, Hamilton BB, Gordon WA, Zasler ND. Characteristics and comparisons of functional assessment indices: Disability Rating Scale, Functional Independence Measure, and Functional Assessment Measure. *J Head Traum Rehabil* 1993; 8: 60–74.
5. Linn RT, Blair RS, Granger CV, Harper DW, O'Hara PA, Maciura E. Does the Functional Assessment Measure (FAM) extend the Functional Independence Measure (FIM) instrument? A Rasch analysis of stroke inpatients. *J Outcome Meas* 1999; 3: 339–359.
6. Turner-Stokes L. Outcome measures for in-patient neurorehabilitation settings – a commentary. *Neuropsychol Rehabil* 1999; 9: 329–343.
7. Turner-Stokes L, Nyein K, Turner-Stokes T, Gatehouse C. The UK FIM+FAM: development and evaluation. *Clin Rehabil* 1999; 13: 277–287.
8. UKROC: UK Rehabilitation Outcomes Collaborative London 2010. Available from: <https://www.kcl.ac.uk/cicelysaunders/research/studies/uk-roc/index> [accessed 2019 Jun 28].
9. Rasch G. Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research). Chicago: The University of Chicago Press; 1960.
10. Andrich D. Rasch models for measurement. Quantitative applications in the social sciences. Andrich D, editor. London: Sage Publications; 1988.
11. Bond TG, Fox JC. Applying the Rasch model. Fundamental measurement in the human sciences. 2nd edn. Toledo: University of Toledo; 2007.
12. 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.
13. Turner-Stokes L, Siegert RJ. A comprehensive psychometric evaluation of the UK FIM + FAM. *Disabil Rehabil* 2013; 35: 1885–1895.
14. Hawley CA, Taylor R, Hellawell DJ, Pentland B. Use of the functional assessment measure (FIM+FAM) in head injury rehabilitation: a psychometric analysis. *J Neurol Neurosurg Psychiatr* 1999; 67: 749–754.
15. Tesio L, Cantagallo A. The functional assessment measure (FAM) in closed traumatic brain injury outpatients: a Rasch-based psychometric study. *J Outcome Meas* 1998; 2: 79–96.
16. 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.
17. Medvedev O, Turner-Stokes L, Siegert RJ. Rasch analysis of the UK Functional Assessment Measure in patients with complex disability after stroke. *J Rehabil Med* 2018; 50: 420–428.
18. Specialist neuro-rehabilitation services: providing for patients with complex rehabilitation needs. London: British Society of Rehabilitation Medicine 2010 (Updated 2015). <https://www.bsrm.org.uk/downloads/specialised-neurorehabilitation-service-standards--7-30-4-2015-pcatv2-forweb-11-5-16-annexe2updatedmay2019.pdf> [accessed 2019 Jun 28].
19. Turner-Stokes L, Vanderstay R, Eagar K, Dredge R, Siegert RJ. Cost-efficient service provision in neurorehabilitation: defining needs, costs and outcomes for people with long-term neurological conditions: Programme grant report (RP-PG-0407-10185). London: National Institute of Health Research, 2015. <https://www.kcl.ac.uk/cicelysaunders/research/studies/uk-roc/index> [Accessed 2019 Jun 28].
20. Turner-Stokes L, Williams H, Bill A, Bassett P, Sephton K. Cost-efficiency of specialist inpatient rehabilitation for working-aged adults with complex neurological disabilities: a multicentre cohort analysis of a national clinical data set. *BMJ Open* 2016; 6: e010238.
21. Hooper D, Coughlan J, Mullen MR. Structural Equation modelling: guidelines for determining model fit. *Electronic J Business Res Meth* 2008; 6: 53–60.
22. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equat Modell* 1999; 6: 1–55.
23. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2030. Perth: RUMM Laboratory; 2010.
24. 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: 127–132.
25. Müller M, Kreiner S. Item fit statistics in common software for Rasch analysis Copenhagen, Denmark: : Department of Biostatistics, University of Copenhagen; 2015. Available from: https://ifsv.sund.ku.dk/biostat/annualreport/images/2/2f/Research_Report_15-06.pdf [accessed 2019 Jun 28].
26. Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch Unidimensional Measurement Model (Rumm) program in health outcome measurement. *J Appl Meas* 2016; 17: 416–431.
27. Holland P, Wainer H. Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum; 1993.
28. Turner-Stokes L, Tonge P, Nyein K, Hunter M, Nielson S, Robinson I. The Northwick Park Dependency Score (NPDS): a measure of nursing dependency in rehabilitation. *Clin Rehabil* 1998; 12: 304–318.
29. Smith EVJ. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002; 3: 205–231.
30. Fan X, Thompson B, Wang L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Struct Equat Model* 1999; 6: 56–83.
31. Feinstein AR. Clinimetric Perspectives. *J Chron Dis* 1987; 40: 635–640.
32. Medvedev ON, Krägeloh CU, Narayanan A, Siegert RJ. Measuring mindfulness: applying generalizability theory to distinguish between state and trait. *Mindfulness* 2017; 8: 1036–1046.