J Rehabil Med 2018; 50: 420–428

ORIGINAL REPORT



# RASCH ANALYSIS OF THE UK FUNCTIONAL ASSESSMENT MEASURE IN PATIENTS WITH COMPLEX DISABILITY AFTER STROKE

Oleg N. MEDVEDEV, PhD[1], Lynne TURNER-STOKES, DM, FRCP[2,3], Stephen ASHFORD, PhD[2,3] and Richard J. SIEGERT, PhD[4]
*From the [1]University of Auckland, School of Medicine, Centre for Medical and Health Sciences Education, Auckland, New Zealand, [2]King's College London, Cicely Saunders Institute, School of Nursing, Midwifery and Palliative Care, Department of Palliative Care Policy and Rehabilitation, [3]Northwick Park Hospital, London North West Healthcare NHS University Trust, London, UK, and [4]Auckland University of Technology, School of Public Health & Psychosocial Studies and School of Clinical Sciences, Auckland, New Zealand*

***Objectives:*** **To determine whether the UK Functional Assessment Measure (UK FIM+FAM) fits the Rasch model in stroke patients with complex disability and, if so, to derive a conversion table of Rasch-transformed interval level scores.**

***Methods:*** **The sample included a UK multicentre cohort of 1,318 patients admitted for specialist rehabilitation following a stroke. Rasch analysis was conducted for the 30-item scale including 3 domains of items measuring physical, communication and psychosocial functions. The fit of items to the Rasch model was examined using 3 different analytical approaches referred to as "pathways".**

***Results:*** **The best fit was achieved in the pathway where responses from motor, communication and psychosocial domains were summarized into 3 super-items and where some items were split because of differential item functioning (DIF) relative to left and right hemisphere location ($\chi^2$ (10) = 14.48, $p$ = 0.15). Re-scoring of items showing disordered thresholds did not significantly improve the overall model fit.**

***Conclusion:*** **The UK FIM+FAM with domain super-items satisfies expectations of the unidimensional Rasch model without the need for re-scoring. A conversion table was produced to convert the total scale scores into interval-level data based on person estimates of the Rasch model. The clinical benefits of interval-transformed scores require further evaluation.**

*Key words:* patient; physiopathology; psychometrics; Rasch; functional assessment.

Accepted Jan 12, 2018; Epub ahead of print Feb 28, 2018

J Rehabil Med 2018; 50: 420–428

Correspondence address: Oleg Medvedev, Centre for Medical and Health Sciences Education, School of Medicine, University of Auckland, Rm 12.025, Bldg 599, 2 Park Rd, Grafton, Auckland 1142, New Zealand. E-mail: o.medvedev@auckland.ac.nz

The Functional Independence Measure is one of the most widely used outcome measures for rehabilitation worldwide, comprising 13 "motor" and 5 "cognitive" items (1, 2). The Functional Assessment Measure was originally developed in the US as an extension of the FIM in the mid-1990s (3, 4), adding a further 12 items to extend its coverage of cognitive and psychosocial function, for use in patients with more complex disabilities following acquired brain injury. Adapted for use in the UK, the UK FIM+FAM was published in 1999 (5). It consists of a 30-item scale encompassing physical, cognitive, communicative and psychosocial function. An optional add-on module addresses extended activities of daily living (6), designed primarily for use in the community. The UK Rehabilitation Outcomes Collaborative (UKROC) provides the national clinical database collating outcomes for all tertiary specialized (Level 1) and local specialist (Level 2) in-patient rehabilitation services in England, and the UK FIM+FAM is now the principal outcome measure within the dataset (7, 8).

The psychometric properties of the 30-item UK FIM+FAM have previously been examined in a general neuro-rehabilitation cohort using exploratory and confirmatory factor analysis (EFA and CFA) and Mokken analysis (a non-parametric technique based on item response theory) (9). These analyses indicated 2 distinct domains: motor (16 items) and cognitive (14 items), the latter dividing into a 5-item communicative and 9-item psychosocial component. This yielded an overall factor structure of 3 subscales (physical, communication and psychosocial), each with a Cronbach's alpha >0.90 and Cohen's d effect sizes ranging from 0.86 to 1.29 between admission and discharge. A subsequent EFA and CFA in stroke patients (10) demonstrated the same 3-factor structure accounted for 69% of the total variance and also identified the anticipated score differences related to hemispheric location of the stroke. The scale was considered to be valid, reliable and responsive to changes occurring in this study population, as well as sensitive to differences that resonate with clinical experience. However, psychometric properties of the UK FIM+FAM have not been tested using the Rasch model (11, 12), which is warranted given its distinct advantages over other more traditional psychometric methods (13, 14).

The Rasch model (11 12), is a robust statistical model that has been applied in numerous psychometric studies to examine and enhance the measurement properties of scales at both the group and individual level (13–18). There are more than 50 published studies that explore how well FIM data conform to the Rasch model including the variety of solutions obtained for the FIM scale, which were tested with and without

doi: 10.2340/16501977-2324

re-ordering of disordered response categories (18). Two previous studies have explored the benefits of Rasch transformation of the original US version of the FIM+FAM in patients following stroke (19) and traumatic brain injury (20). However, as yet there have been no published Rasch analyses of the UK FIM+FAM in any population. The aim of this paper was to assess the psychometric properties of the UK FIM+FAM in stroke patients with complex disability using Rasch methodology and to produce a conversion table to convert ordinal to interval quality data.

## METHODS

### *Data source, sampling and measure*

*Data source.* The data source was the UKROC database, which was initially set up by a National Institute for Health Research Programme Grant (7, 8). It is now commissioned by NHS England to provide the national clinical 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 in-patient rehabilitation. Since April 2013, reporting of the full UKROC dataset is a mandated requirement for commissioning of all Level 1 and 2 specialist rehabilitation services. However, reporting was voluntary until that date, so not all services routinely reported UK FIM+FAM data. Within these Level 1/2 services, which have a mean (standard deviation (SD)) length of stay of approximately 80 days (SD 60), 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.

The programme is registered with the NIHR Comprehensive Local Research Network: ID number 6352.

*Sampling.* We extracted the cohort of all 1318 stroke patients consecutively admitted to the 58 Level 1/2 specialist rehabilitation centres in England that submitted data to the UKROC database between 1 January 2010 and 30 May 2013, for whom a complete UK FIM+FAM score was available at both admission and discharge from the unit.

- FIM+FAM scores are expected to be lower on admission and higher at discharge from rehabilitation. To ensure that the data represented the full range of the scale, the complete sample of $n = 1{,}318$ included approximately equal numbers of admission and discharge patients.
- In order not to violate the Rasch assumption of local independence between observations (i.e. to prevent the same patient contributing 2 entries in the data) we included only 1 time-point, i.e. admission or discharge, for each patient (21). Fig. 1 summarizes the process of extraction and analysis.

*Measure.* Within the UK FIM+FAM, each of the 30 items is scored on the same 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, while for categories 1–4 the assessment is based on the amount of help required, e.g. the percentage of task performed by patient. The UKROC software automatically produces a "FIM+FAM-Splat" or radar chart, presenting a visual impression of change at item level. This may be used to describe change in individual scores, or median scores for a population, in a format that is clinically interpretable by rehabilitation professionals. By way of example, Fig. 2 shows a composite FIM+FAM-Splat for median admission and discharge scores within this dataset.

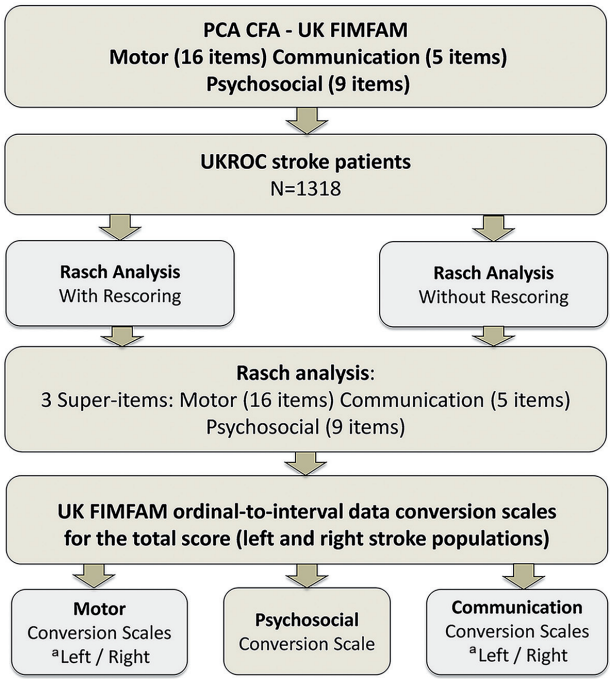


**Fig. 1.** Flow chart of the study extraction and analysis. [a]Left/right stroke DIF by stroke location led to different conversion scales for left and right stroke). UKROC: UK Rehabilitation Outcome Collaborative database; UK FIMFAM: UK Functional Independence Measure and Functional Assessment Measure.

Summing the item scores gives a total range from 30 to 210, where a maximum score of 210 indicates total independence. The 7-category structure implies, in Rasch terms, that each item has 6 possible thresholds or points between 2 response categories where either response is equally probable (i.e.1–2, 2–3, etc.). The original scores format 1–7 was re-coded into 0–6 format for the purpose of analysis as required by the partial credit Rasch model (18).

### *Psychometric analysis of the UK FIM+FAM*

There is now an extensive literature providing guidance methodology for Rasch analysis. Lundgren-Nilsson & Tennant (18) have examined specifically the literature applying the Rasch model to the FIM™ describing how the approach has evolved over 2 decades and making recommendations to improve the rigor of future analyses. During this analysis, we followed their suggestions using different analytical strategies referred to as "pathways" to address issues of local dependence, DIF and disordered thresholds without (if at all possible) removing items to maintain the clinical integrity of the instrument.

Like Lundgren-Nilsson & Tennant and Lundgren-Nilsson et al. (18, 22), we distinguish between local response dependence and local trait dependence (see also in Discussion). Problems due to local response dependence may be dealt with by construction of super-items summarizing item scores from the set of locally dependent items. If the subsequent analysis accepts the distributions of these super-items similar to partial credit items depending

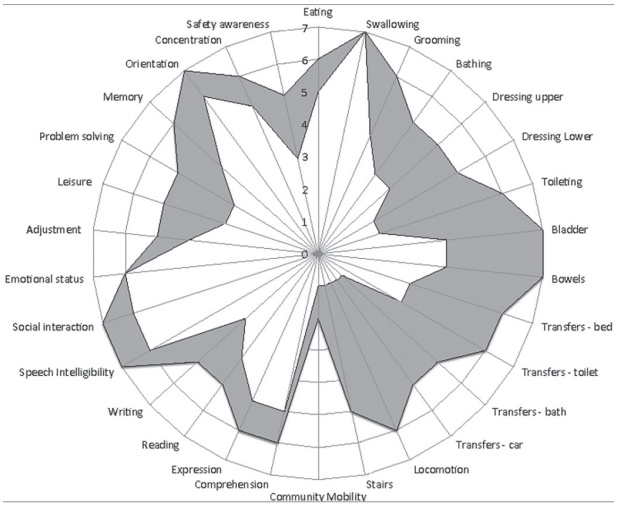


**Fig. 2.** Composite FIM+FAM-splats of the median admission and discharge scores for each item within this dataset. The radar chart (or "FIM+FAM splat") provides a graphic representation of the disability profile from the FIM+FAM data. The 30-scale items are arranged as spokes of a wheel. Scoring levels from 1 (total dependence) to 7 (total independence) run from the centre outwards. Thus, a perfect score would be demonstrated as a large circle. These composite radar charts illustrate the median admission and discharge scores within this dataset. The yellow-shaded portion represents the median admission scores and the blue-shaded area represents the difference between median scores on admission and discharge.

on the same latent variable, it may be taken as evidence against local trait dependence, because local trait dependence due to multidimensionality cannot generate super-items (22).

Prior to commencing our analysis we applied item-trait interaction tests (18) in RUMM2030, which indicated that assumptions of the polytomous Rating scale model did not hold and thus supported appropriateness of the unrestricted Partial Credit Model for Rasch analysis (12, 18). Rasch model fit statistics used to determine fit to the Rasch model included the item-trait interaction $\chi^2$ (overall and individual items), the DIF test and correlations between response residuals (18). Standard errors of the estimates of person parameters were applied to estimate measurement error. The Person Separation Index (PSI) is a measure of scale reliability and represents a function of the variance of the person parameters and the standard error of measurement. PSI values above 0.7 are required for group use and above 0.8 for individual assessment. A residual correlation above 0.2 with reference to the mean of all residual correlations is considered as an indicator of local dependency (23). The first analytical pathway involved the initial Rasch analysis of all 30 items to assess the overall and individual item fit. The second analytical pathway used "super-items" to address local dependency issues without re-scoring disordered thresholds. A disordered threshold occurs when people higher in the ability or construct being measured (in this case *independence*) do not consistently obtain correspondingly higher response options (i.e. 1, 2, 3–7 ) for an item. However, evidence of a disordered threshold can appear for reasons other than the order of the categories. In particular, local response dependence may create evidence of disordered threshold because the dependence distorts the distribution of the separate items. The third analytical pathway involved re-scoring of significantly disordered thresholds for individual items prior to further analysis.

In Rasch analysis, disordered thresholds are corrected by collapsing adjacent response categories. We re-scored items with significantly disordered thresholds by collapsing adjacent categories in a meaningful way (e.g. "total" and "maximal assistance"; "supervision/set-up" and "modified independence").

In the subsequent pathways, we tested for item bias across important person factors such as age group (0–44, 45–54, 55–64, 65–74, 75+ years), sex, ethnicity, type of stroke (haemorrhagic, infarct, sub-arachnoid and other), stroke location (left or right hemisphere) and time-point (admission or discharge). Andrich & Hagquist (24) introduced the concept of "artificial DIF" that may result when real DIF in 1 item favouring 1 group induces artificial DIF favouring the other group in other items. They provided recommendations to deal with DIF issues. We have used these recommendations and, if DIF was found, we resolved it sequentially to differentiate between real and artificial DIF. If uniform DIF for a specific person factor was identified in 1 or more items, the item displaying the strongest DIF effect was split first to allow variation by the corresponding factor and DIF analysis was repeated for other items (24).

As it was desirable to keep the original structure of the UK FIM+FAM scale, item removal was considered only as a last resort to improve the fit. The items at risk of deletion were those exhibiting significant misfit, i.e. excessive item fit-residual values outside ± 2.5 range and a $p$-value significant at the 0.05 level, with a Bonferroni adjustment for multiple tests.

Unidimensionality was tested using principal components analysis (PCA) of the residuals and the equating $t$-test. Unidimensionality of the scale is confirmed if significant $t$-test comparisons do not exceed 5%, or if the lower bound of a binominal confidence interval computed for the number of significant $t$-tests overlaps the 5% cut-off point (25). We followed the recently published guidelines and recommendations for reporting Rasch analysis (26).

### *Statistical analysis and software*

Descriptive analysis was carried out using the IBM SPSS v. 22 software. Rasch analysis was performed using RUMM2030 software (27). The overall item-trait interaction $\chi^2$ and $p$-values generated by RUMM2030 software may be misleading in larger samples (28). Hagell & Westergren (29) suggested that estimation of type I errors is only accurate if $n < 500$. Therefore, we used a random sample of recommended size $n = 320$ (30) to compute the overall $\chi^2$ fit statistics and presented it together with the values obtained for the full sample for comparisons. A significance value of 0.05 was used throughout.

## RESULTS

Within our clinical sample of 1,318 cases, the mean age was 58.91 (SD 15.59) years, range 13–100, 29 participants had missing age data. From this sample we extracted a random sample ($n = 320$) for the purpose of Rasch analysis that displayed comparable demographic characteristics (Table I).

Table II presents the overall Rasch model fit statistics for all 3 analytical pathways described above, including the item-trait interaction $\chi^2$ and $p$-values values for the random sample ($n = 320$) and for the full sample (in parentheses). Table III presents the Rasch model results for each individual item, along with the

**Table I.** The UK Rehabilitation Outcomes Collaborative (UKROC): stroke population sample characteristics

| Demographic characteristics | UKROC Study sample $n = 1,318$ | Random sample[a] $n = 320$ |
|---|---|---|
| Age, *n* (%) | | |
| < 44 years | 220 (16.7) | 50 (15.6) |
| 45–54 years | 293 (22.2) | 74 (23.1) |
| 55–64 years | 298 (22.6) | 66 (20.6) |
| 65–74 years | 250 (19.0) | 54 (16.9) |
| ≥ 74 years | 231 (17.5) | 68 (21.3) |
| Unknown | 26 (2.0) | 8 (2.5) |
| Male, *n* (%) | 752 (57.1) | 189 (59.1) |
| Ethnicity, *n* (%) | | |
| White | 951 (72.2) | 227 (70.9) |
| Asian/Asian British | 98 (7.4) | 21 (6.6) |
| Black/Black British | 110 (8.3) | 29 (9.1) |
| Other | 41 (3.1) | 10 (3.2) |
| Unknown | 118 (8.9) | 33 (10.3) |
| Length of stay, days, mean (SD) | 77.7 (57.3) | 78.9 (52.6) |
| Diagnosis localization, *n* (%) | | |
| Right hemisphere | 638 (48.4) | 159 (49.7) |
| Left hemisphere | 680 (51.6) | 161 (50.3) |
| Diagnosis subcategory, *n* (%) | | |
| Haemorrhagic | 386 (29.3) | 93 (29.1) |
| Infarct | 707 (53.6) | 174 (54.4) |
| Sub-arachnoid | 136 (10.3) | 32 (10.0) |
| Other | 89 (6.8) | 21 (6.6) |

[a]Random sample extracted from the dataset ($n = 1,318$) derived across admission and discharge values so that each patient is only in the dataset once, but both time-points are equally represented.

frequency distribution of responses for each of the 7 scoring categories within the 30 items. There are no categories endorsed by less than 20 responses. We identified 15 persons with extreme locations above 4 logits and negative fit residuals below –5 that may significantly affect the estimates (26) and presented data both with and without these 15 extreme scores for comparison.

### *Analytical pathway 1: Initial analysis of the full 30-item scale*

The initial analysis including all 30 items showed equally good reliability with and without extreme persons (PSI = 0.95–0.96), but misfit at both individual item and the overall level with significant item-trait interaction. Table III shows significant misfit for 18 out of 30 items. At this stage the residual correlation matrix was examined and it displayed local dependencies between 3 groups of items that mirrored our previously reported results of factor analysis (9, 10), i.e. Motor (16 items), Communication (5 items) and Psychosocial (9 items) function). For the next stage of the analysis, the 30 items were combined into 3 super-items representing motor, communication and psychosocial function.

### *Analytical pathway 2: Super-items analysis without re-scoring*

*Pathway 2A.* Table II shows that super-items analysis without re-scoring produced satisfactory overall model fit with and without extremes in the random sample ($n = 320$). Even though, $\chi^2$ values were relatively small in the full sample, *p*-values indicated errors that were not observed in the random sample of sufficient size. Unidimensionality was confirmed in the full sample with only 1.21% of *t*-tests significant (see Table II). However, reliability of analysis 2B with extreme persons was below the acceptable level. DIF analysis indicated significant uniform DIF for the Motor ($F(1,1301) = 93.05$, $p < 0.001$) and Communication subtests ($F(1,1301) = 353.25$, $p < 0.001$) by stroke localization without extremes that was then replicated with extremes included in the sample, but no other DIF was identified (Fig. 3).

*Pathway 2C and D.* Communication super-item was split for DIF by localization first because it showed stronger uniform DIF effect. This did not resolve DIF in the motor super-item suggesting real DIF by localization. Therefore, motor super-item was also split for DIF by localization. When the motor and com-

**Table II.** The UK Functional Assessment Measure (UK FIM+FAM): Rasch model summary statistics (overall fit of the scale)

| UK FIM FAM Rasch model | Item –trait interaction[a] $\chi^2$/DF | *p*-value | PSI | Unidimensional (Sig. *t*-test %) | Local dependency |
|---|---|---|---|---|---|
| *Pathway 1: All 30 items* | | | | | |
| Analysis 1A (no extremes) | 540.87/120 (500.23/120) | 0.00 (0.00) | 0.96 | No (> 40) | Yes |
| Analysis 1B (with extremes) | 540.87/120 (500.23/120) | 0.00 (0.00) | 0.95 | No (> 40) | Yes |
| *Pathway 2: Three super-items no re-scoring* | | | | | |
| Analysis 2A (no extremes) | 20.81/12 (84.04/12) | 0.05 (0.00) | 0.79 | Yes (1.21) | No |
| Analysis 2B (with extremes) | 20.81/12 (54.79/12) | 0.05 (0.00) | < 0.50 | Yes (1.21) | No |
| Analysis 2C (DIF split no extremes) | 14.48/10 (76.86/10) | 0.15 (0.00) | 0.80 | Yes (1.21) | No |
| Analysis 2D (DIF split with extremes) | 14.48/10 (76.85/10) | 0.15 (0.00) | 0.76 | Yes (1.21) | No |
| *Pathway 3: Three super-items with re-scoring* | | | | | |
| Analysis 3A (without extremes) | 18.65/12 (76.80/12) | 0.10 (0.00) | 0.79 | Yes (1.46) | Yes (2 & 3)[b] |
| Analysis 3B (with extremes) | 18.65/12 (76.80/12) | 0.10 (0.00) | 0.78 | Yes (1.52) | Yes (2 & 3)[b] |

[a]Item –Trait Interaction $\chi$ square and *p*-values are based on the random sample $n = 320$ and values in brackets, Person Separation Index (PSI), unidimensionality and local dependency tests are estimated with the full sample $n = 1,318$; [b]2 = Super-item of Communication domain; 3 = Super-item of Psychosocial domain.

**Table III.** Frequency distribution of responses and Rasch model fit statistics for the UK FIM+FAM items (Pathway 1, Analysis 1B), and domain super-items split by localization without re-scoring (Pathway 2, Analysis 2C), $n = 1,318$

| | | | | | Frequency distribution across scoring categories | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Description | Location | Fit residual | $\chi^2$ | Cat 1 | Cat 2 | Cat 3 | Cat 4 | Cat 5 | Cat 6 | Cat 7 |
| *Pathway 1: All 30 items initial analysis* | | | | | | | | | | | |
| 1 | Eating[a] | −0.59 | −1.33 | 32.17 | 88 | 29 | 31 | 68 | 425 | 195 | 467 |
| 2 | Swallowing[a] | −1.03 | 1.28 | 19.66 | 45 | 24 | 48 | 28 | 114 | 102 | 942 |
| 3 | Grooming* | −0.22 | **−8.48** | **123.45** | 107 | 105 | 132 | 159 | 330 | 179 | 291 |
| 4 | Bathing* | 0.23 | **−9.43** | **125.30** | 194 | 168 | 211 | 206 | 215 | 136 | 173 |
| 5 | Dressing – upper* | 0.00 | **−8.46** | **97.91** | 162 | 151 | 200 | 189 | 185 | 151 | 265 |
| 6 | Dressing – lower* | 0.34 | **−9.82** | **106.61** | 299 | 226 | 157 | 158 | 130 | 116 | 217 |
| 7 | Toileting*[a] | 0.15 | **−6.96** | **77.56** | 328 | 164 | 106 | 114 | 93 | 157 | 341 |
| 8 | Bladder*[a] | 0.07 | **−7.72** | 66.67 | 290 | 76 | 177 | 111 | 151 | 155 | 343 |
| 9 | Bowel*[a] | 0.12 | **−8.21** | 66.90 | 320 | 68 | 165 | 109 | 137 | 188 | 316 |
| 10 | Bed transfers[a] | 0.42 | −2.12 | 13.55 | 486 | 66 | 128 | 96 | 154 | 165 | 208 |
| 11 | Toilet transfers[a] | 0.53 | −1.19 | 6.51 | 644 | 30 | 76 | 99 | 128 | 100 | 226 |
| 12 | Bath transfers | 0.75 | −1.86 | 7.15 | 764 | 22 | 35 | 48 | 125 | 173 | 136 |
| 13 | Car transfers[a] | 1.18 | −1.94 | 32.56 | 760 | 127 | 166 | 42 | 92 | 68 | 48 |
| 14 | Locomotion[a] | −0.08 | 2.01 | 19.22 | 287 | 94 | 78 | 86 | 99 | 141 | 518 |
| 15 | Stairs*[a] | −0.12 | **−3.94** | 20.78 | 282 | 88 | 73 | 59 | 100 | 142 | 559 |
| 16 | Community mobility[a] | 0.44 | −1.44 | 10.89 | 540 | 64 | 32 | 51 | 183 | 230 | 203 |
| 17 | Comprehension* | −0.40 | **4.11** | 24.82 | 68 | 91 | 103 | 124 | 250 | 322 | 345 |
| 18 | Expression* | −0.17 | **6.95** | **108.72** | 142 | 132 | 106 | 96 | 181 | 261 | 385 |
| 19 | Reading*[a] | 0.06 | **5.24** | 62.43 | 262 | 70 | 76 | 127 | 239 | 225 | 304 |
| 20 | Writing*[a] | 0.28 | **8.44** | **201.70** | 390 | 106 | 120 | 97 | 160 | 168 | 262 |
| 21 | Speech intelligibility* | −0.51 | **8.49** | **114.31** | 78 | 72 | 99 | 79 | 141 | 202 | 632 |
| 22 | Social Interaction* | −0.70 | **5.85** | **61.25** | 49 | 68 | 64 | 66 | 176 | 290 | 590 |
| 23 | Emotional Status* | −0.33 | **12.20** | **247.54** | 97 | 104 | 69 | 61 | 176 | 352 | 444 |
| 24 | Adjustment to limitations* | −0.06 | **4.66** | **30.93** | 114 | 176 | 204 | 115 | 221 | 263 | 210 |
| 25 | Use of leisure time[a] | 0.20 | −0.90 | 22.07 | 185 | 195 | 182 | 123 | 123 | 378 | 117 |
| 26 | Problem solving* | 0.34 | −2.35 | **40.08** | 215 | 226 | 116 | 147 | 266 | 211 | 122 |
| 27 | Memory | −0.07 | **4.97** | **63.59** | 157 | 148 | 159 | 153 | 157 | 213 | 316 |
| 28 | Orientation*[a] | −0.45 | **2.59** | 24.28 | 115 | 77 | 105 | 109 | 101 | 154 | 642 |
| 29 | Concentration | −0.41 | 1.92 | 19.76 | 73 | 112 | 116 | 147 | 244 | 232 | 379 |
| 30 | Safety awareness[a] | 0.04 | 1.89 | 13.93 | 86 | 342 | 218 | 163 | 118 | 182 | 194 |
| *Pathway 2: Three super-items – no re-scoring Analysis 2C (DIF split without extremes)* | | | | | | | | | | | |
| 1 | Motor Left | 0.03 | −2.54 | 12.70 | | | | | | | |
| 2 | Motor Right | 0.07 | −2.32 | 13.71 | | | | | | | |
| 3 | Communication left | 0.02 | 1.45 | 21.07 | | | | | | | |
| 4 | Communication right | −0.11 | 1.57 | 26.11 | | | | | | | |
| 5 | Psychosocial | −0.02 | 2.54 | 30.16 | | | | | | | |

*Significant misfit to the Rasch model ($p < 0.05$, Bonferroni adjusted).
[a]Denotes items with significantly disordered thresholds ($p < 0.05$).
Bold numbers indicate fit parameters associated with significant misfit to the Rasch model.

munication subtests were split by localization (left/right) to control for DIF, this produced the best model fit with and an improved PSI of 0.80 (Table II). At this stage, the scale was strictly unidimensional and there were no locally dependent or significantly misfitting super-items identified (Table II and III, Pathway 2, Analysis 2C). This analysis was replicated with extreme persons, resulting in equally good fit but lower reliability (PSI = 0.76).

### *Analytical pathway 3: Super-items analysis with re-scoring*

*Pathway 3A and B.* Applying the third analytical pathway (with re-scoring), significantly disordered thresholds were identified in 15 out of 30 items. Table III indicates the items with significantly disordered thresholds. Notably, of the 15 items with disordered thresholds only 3 items (number 8 (Bed transfers), number 9 (Toilet transfers), and number 20 (Writing)) are misfitting. All 15 items with disordered thresholds were re-scored before the analysis continued. After re-scoring, the items showed similar patterns of local dependency and were combined into motor, communication and psychosocial subtests. The resultant fit indices were comparable to those achieved without re-scoring in both the analyses with and without extremes, but the reliability was higher in the analysis with extremes when disordered items were re-scored. However, local dependency between the 2 super-items *communication and psychosocial* that exceeded the accepted cut-off point of 0.2 was identified. An attempt to combine these super-items into 1 single super-item resulted in a decrease in reliability (PSI = 0.71), below that which was desirable for individual assessment.

Fig. 4 presents the item-person threshold distributions of the best solution without re-scoring (Analysis 2C). It can be seen that abilities of the sample are fairly

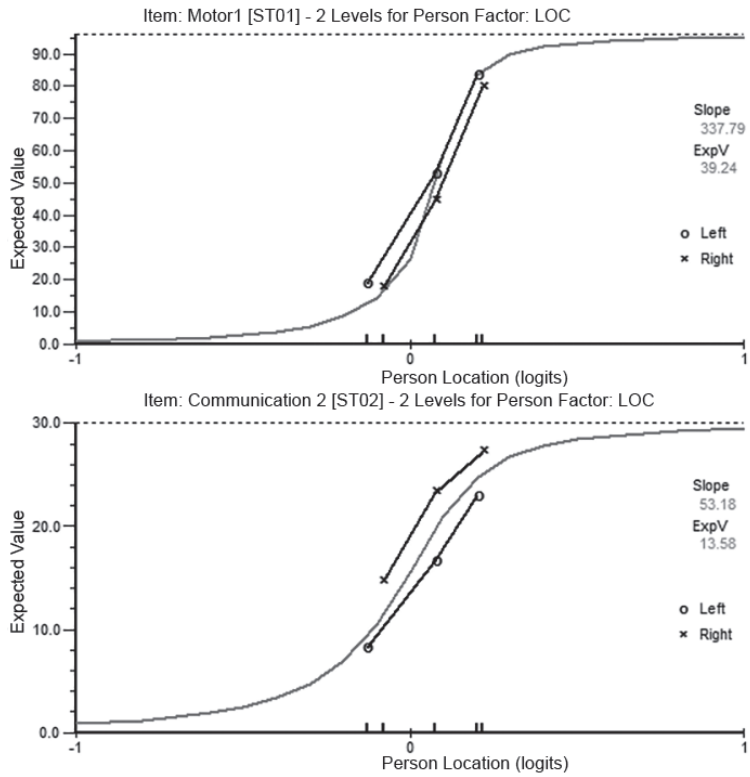


**Fig. 3.** Item characteristic curves (ICC) with uniform DIF by localization for the motor super-item (top panel) and communication super-item (bottom panel). Analysis 2A without extremes.

well targeted by item thresholds without any significant ceiling or floor effects, and person distribution is close to a normal distribution. Therefore, the scale version without re-scoring that achieved the best model fit (Analysis 2C) was used to generate ordinal-to-interval conversion tables. Standard errors of measurement for raw scores of 50, 100, 150, and 175 (left stroke) were 4.47, 2.95, 3.21 and 3.39, respectively, and similarly small values were estimated for right hemisphere stroke patients. Table IV contains ordinal-to-interval conversion scores estimated from the analysis without re-scoring disordered thresholds and not including extreme persons.

The left and right location scales were strongly correlated ($r=0.99, p<0.001$), but paired *t*-test comparisons demonstrated significant differences between the 2 scales ($t(180)=-4.22; p<0.001$). A scatter plot (Fig. 5), however, shows that the differences between left and right scores is actually very small (see discussion).

## DISCUSSION

The study presented here represents the first Rasch analysis of the UK FIM+FAM, which is the primary outcome measure within the UKROC national clinical dataset for all specialist rehabilitation services in the UK treating patients with complex disabilities.

The best fit to the Rasch measurement model was achieved when 3 groups of locally-dependent items were treated as super-items, which provides strong evidence of unidimensionality for the UK FIM+FAM. Preliminary results of factor analysis (9, 10) indicated that 3 domains Motor, Communication and Psychosocial represent different factors because items of each domain share common variance. However, shared variance may be evident for 2 different reasons: multidimensionality due to "trait dependence" (i.e. the tool genuinely measures different constructs), or "response dependence", where the response to 1 item influences responses to other related items (18). Multidimensional measures representing different traits typically fail to fit the strict criteria of the unidimensional Rasch model, which complies with the principles of fundamental measurement formulated by Thurstone (31), such as unidimensionality, sample invariance and a consistent unit of measurement across the scale continuum. The findings from this analysis indicate that the UK FIM+FAM

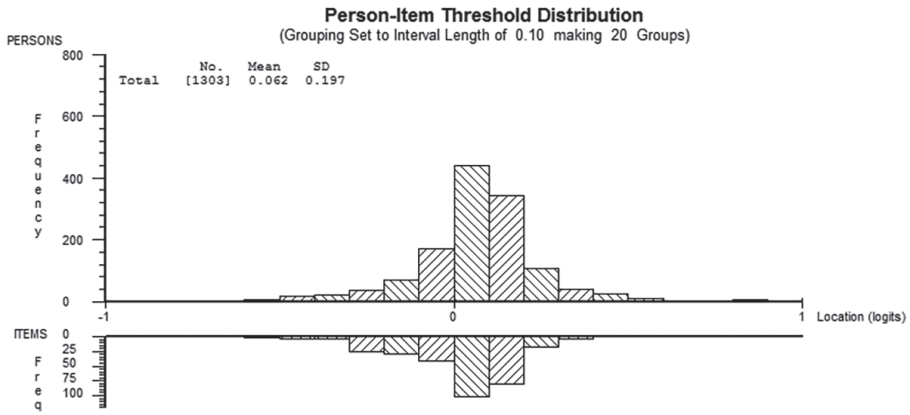


**Fig. 4.** Person-item threshold distributions for the final solution without re-scoring (*top panel*) and with re-scoring (*bottom panel*) for the left and right stroke populations.

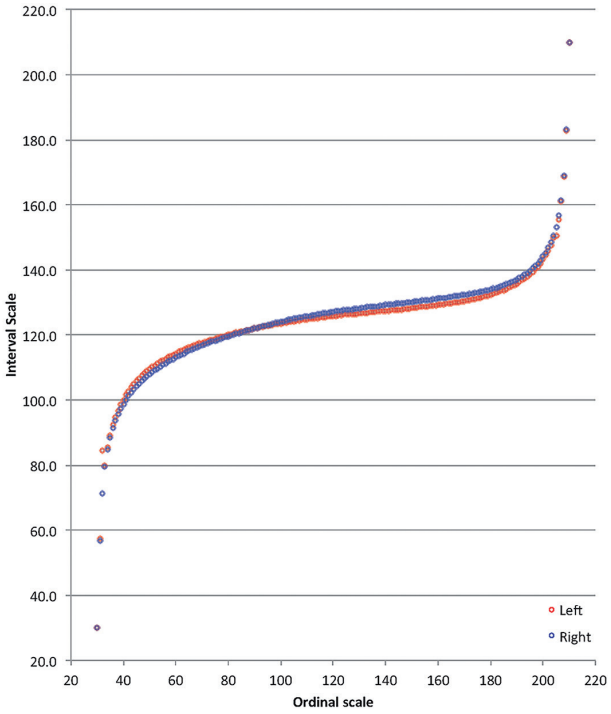


**Fig. 5.** Scatter plot of Rasch interval-level scores (y-axis) as a function of ordinal scale scores (x-axis).

satisfies the Rasch model without the need to re-score disordered thresholds in a random sample of stroke patients. This allowed for a very simple conversion from raw scores to an interval metric for the whole scale.

Two previous studies have explored Rasch analysis of the original US version (USFIM+FAM), using the WINSTEPS software (19, 20):

- Linn et al. (19) also reported a number of misfitting items, but they were principally interested in whether the FAM solved the problem of ceiling effects in the FIM. This has limited relevance to the present study as the UK FIM+FAM has dealt with ceiling effects in a different way, by providing a separate module addressing extended activities of daily living (6) as well as a related scale of workability (32).
- Hawley et al. (20) examined the US FIM+FAM in a cohort of 652 patients with traumatic brain injury (TBI). They used a principal component analysis to identify 2 separate dimensions (Motor-16 items and Cognitive 14 items), which conformed only partially to the Rasch model. As they point out, the imperfect fit is hardly surprising given the heterogeneity of a typical brain injury sample and the diverse nature of the items captured within the FIM+FAM. It does not necessarily indicate that the scale is fundamentally flawed in a clinical sense. The question that arises, however, is what further division of subscales is necessary to improve the fit, and these considerations may apply equally to a complex stroke population.

These early studies reported the goodness of overall and individual item fit to the Rasch model, but typically went little further. They frequently relied on deleting items to attain satisfactory fit and rarely provided a table to permit the conversion of raw scores to interval level scores in routine clinical practice. A major methodological strength of our study is that we were able to draw upon 21 years of experience in Rasch studies on the FIM, following the methodology described by Lundgren-Nilsson & Tennant and Lundgren-Nilsson et al. (18, 22) to explore how well the UK FIM+FAM fits the Rasch model according to more current analytical techniques. We used a range of steps including the formation of super-items to eliminate local dependency among items to achieve reasonably good fit for the 3 dimensions underpinning the UK FIM+FAM. Importantly, we have recognized the difference between local response dependence and local trait dependence and calculated super-items to address the dependence among items. We were able to do this without deleting any items and also to produce a conversion table for left and right hemisphere strokes, to account for differential item functioning between these 2 groups.

First, it is clinically expected that a left hemisphere stroke is generally less associated with motor impairments compared with right hemisphere strokes, which is consistent with our uniform DIF-findings for the motor super-item. On the other hand, left hemisphere strokes are more frequently linked to impairments in communication compared with the right hemisphere that is again consistent with our results for the Communication super-item. The DIF split does not affect the validity of the measure as evidenced by a strong correlation between conversion interval scores for left and right stroke population and reflected by the scatterplot (Fig. 5), the actual difference is very small and unlikely to be clinically important.

The chief advantage of measures that conform to the Rasch model is that their data can be analysed with parametric statistics rather than relying on non-parametric statistics lending greater statistical power and precision. Whilst the use of interval level scales has some clear advantages for the generation of robust metrics for the purpose of research, further work is necessary to explore the impact and benefits of transformed scores in the clinical setting. We recognize that, despite the many conversion tables that have been produced for the FIM in different contexts (17), the uptake of these by clinicians has been limited because the ordinal scores within each item are interpretable at a clinical level and are widely used as an aid to clinical reporting and decision-making. The FAM splat is particularly valued by UK clinicians in this context, and

**Table IV.** The UK FIM+FAM conversion scale: the raw scores and corresponding Rasch interval scores accounting for left and right strokes differential item functioning

| Raw score | Interval | | Raw score | Interval | | Raw score | Interval | | Raw score | Interval | | Raw score | Interval | | Raw score | Interval | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Left | Right | | Left | Right | | Left | Right | | Left | Right | | Left | Right | | Left | Right |
| 30 | 30.0 | 30.0 | 61 | 114.7 | 113.5 | 92 | 122.4 | 122.4 | 123 | 126.1 | 127.5 | 154 | 128.7 | 130.6 | 185 | 133.8 | 135.3 |
| 31 | 57.3 | 56.8 | 62 | 115.1 | 113.9 | 93 | 122.6 | 122.7 | 124 | 126.3 | 127.6 | 155 | 128.8 | 130.7 | 186 | 134.2 | 135.7 |
| 32 | 84.7 | 71.4 | 63 | 115.5 | 114.4 | 94 | 122.7 | 122.9 | 125 | 126.3 | 127.8 | 156 | 128.9 | 130.7 | 187 | 134.6 | 135.9 |
| 33 | 80.0 | 79.4 | 64 | 115.8 | 114.7 | 95 | 122.8 | 123.1 | 126 | 126.4 | 127.8 | 157 | 129.0 | 130.8 | 188 | 134.9 | 136.3 |
| 34 | 85.3 | 84.7 | 65 | 116.1 | 115.1 | 96 | 123.0 | 123.3 | 127 | 126.5 | 127.9 | 158 | 129.0 | 131.0 | 189 | 135.4 | 136.6 |
| 35 | 89.3 | 88.5 | 66 | 116.4 | 115.4 | 97 | 123.2 | 123.4 | 128 | 126.4 | 128.1 | 159 | 129.2 | 131.0 | 190 | 135.8 | 137.1 |
| 36 | 92.3 | 91.4 | 67 | 116.7 | 115.8 | 98 | 123.3 | 123.7 | 129 | 126.6 | 128.2 | 160 | 129.3 | 131.2 | 191 | 136.3 | 137.6 |
| 37 | 94.8 | 93.7 | 68 | 117.1 | 116.2 | 99 | 123.5 | 123.8 | 130 | 126.7 | 128.2 | 161 | 129.4 | 131.3 | 192 | 136.8 | 138.0 |
| 38 | 96.9 | 95.7 | 69 | 117.4 | 116.5 | 100 | 123.6 | 124.1 | 131 | 126.7 | 128.4 | 162 | 129.5 | 131.3 | 193 | 137.4 | 138.6 |
| 39 | 98.6 | 97.4 | 70 | 117.6 | 116.8 | 101 | 123.7 | 124.2 | 132 | 126.9 | 128.5 | 163 | 129.6 | 131.5 | 194 | 138.0 | 139.1 |
| 40 | 100.2 | 98.8 | 71 | 117.9 | 117.1 | 102 | 123.9 | 124.5 | 133 | 126.8 | 128.6 | 164 | 129.8 | 131.6 | 195 | 138.6 | 139.7 |
| 41 | 101.5 | 100.1 | 72 | 118.2 | 117.5 | 103 | 124.0 | 124.6 | 134 | 127.1 | 128.7 | 165 | 129.8 | 131.7 | 196 | 139.4 | 140.5 |
| 42 | 102.8 | 101.3 | 73 | 118.4 | 117.7 | 104 | 124.2 | 124.8 | 135 | 127.1 | 128.7 | 166 | 130.0 | 131.9 | 197 | 140.2 | 141.3 |
| 43 | 103.9 | 102.3 | 74 | 118.6 | 118.0 | 105 | 124.3 | 125.0 | 136 | 127.1 | 128.9 | 167 | 130.1 | 131.9 | 198 | 141.0 | 142.0 |
| 44 | 104.9 | 103.3 | 75 | 118.9 | 118.3 | 106 | 124.4 | 125.1 | 137 | 127.3 | 128.8 | 168 | 130.2 | 132.0 | 199 | 142.0 | 143.0 |
| 45 | 105.8 | 104.2 | 76 | 119.2 | 118.6 | 107 | 124.5 | 125.3 | 138 | 127.3 | 129.1 | 169 | 130.4 | 132.2 | 200 | 143.2 | 144.1 |
| 46 | 106.6 | 105.0 | 77 | 119.4 | 118.9 | 108 | 124.6 | 125.5 | 139 | 127.5 | 129.0 | 170 | 130.5 | 132.4 | 201 | 144.4 | 145.4 |
| 47 | 107.5 | 105.8 | 78 | 119.6 | 119.1 | 109 | 124.7 | 125.6 | 140 | 127.5 | 129.3 | 171 | 130.7 | 132.5 | 202 | 145.9 | 146.8 |
| 48 | 108.2 | 106.5 | 79 | 119.8 | 119.4 | 110 | 124.8 | 125.8 | 141 | 127.6 | 129.3 | 172 | 130.8 | 132.6 | 203 | 147.7 | 148.5 |
| 49 | 108.9 | 107.3 | 80 | 120.1 | 119.6 | 111 | 124.9 | 125.9 | 142 | 127.6 | 129.3 | 173 | 131.0 | 132.7 | 204 | 149.8 | 150.7 |
| 50 | 109.5 | 107.9 | 81 | 120.3 | 119.9 | 112 | 125.1 | 126.1 | 143 | 127.7 | 129.5 | 174 | 131.2 | 132.9 | 205 | 150.6 | 153.2 |
| 51 | 110.1 | 108.6 | 82 | 120.5 | 120.1 | 113 | 125.2 | 126.2 | 144 | 127.7 | 129.7 | 175 | 131.3 | 133.1 | 206 | 155.4 | 156.6 |
| 52 | 110.7 | 109.2 | 83 | 120.7 | 120.4 | 114 | 125.3 | 126.4 | 145 | 127.9 | 129.7 | 176 | 131.5 | 133.2 | 207 | 161.0 | 161.5 |
| 53 | 111.2 | 109.7 | 84 | 120.9 | 120.6 | 115 | 125.4 | 126.5 | 146 | 127.9 | 129.7 | 177 | 131.7 | 133.4 | 208 | 168.6 | 169.0 |
| 54 | 111.7 | 110.2 | 85 | 121.1 | 120.9 | 116 | 125.4 | 126.6 | 147 | 128.0 | 129.8 | 178 | 131.9 | 133.6 | 209 | 183.0 | 183.2 |
| 55 | 112.2 | 110.8 | 86 | 121.3 | 121.1 | 117 | 125.5 | 126.8 | 148 | 128.2 | 129.9 | 179 | 132.2 | 133.8 | 210 | 210.0 | 210.0 |
| 56 | 112.7 | 111.3 | 87 | 121.5 | 121.3 | 118 | 125.7 | 126.9 | 149 | 128.2 | 130.1 | 180 | 132.4 | 134.0 | | | |
| 57 | 113.1 | 111.8 | 88 | 121.6 | 121.5 | 119 | 125.8 | 127.0 | 150 | 128.3 | 130.2 | 181 | 132.6 | 134.3 | | | |
| 58 | 113.6 | 112.3 | 89 | 121.8 | 121.8 | 120 | 125.8 | 127.1 | 151 | 128.5 | 130.3 | 182 | 132.9 | 134.4 | | | |
| 59 | 114.0 | 112.7 | 90 | 122.0 | 122.0 | 121 | 125.9 | 127.3 | 152 | 128.5 | 130.4 | 183 | 133.2 | 134.7 | | | |
| 60 | 114.4 | 113.1 | 91 | 122.2 | 122.2 | 122 | 126.0 | 127.4 | 153 | 128.6 | 130.4 | 184 | 133.5 | 135.0 | | | |

for this reason we would not necessarily recommend using transformed scores at the individual item level, although they may nevertheless prove valuable when presenting summed items in subscale and total scores, particularly if the transformed data prove to be more sensitive (16). However, Fig. 5 demonstrates that the interval level is markedly "flat" in the middle part of the scale, changing by just 40 points (100–140) while the ordinal scale changes by 157 points (40–197). As demonstrated by the FAM splat, this is the part of the scale in which the majority of patients are likely to show change. Thus, while the interval scale may provide more reliable measurement at a statistical level, it may not be responsive to clinical change. The benefits of transformed scores therefore require further evaluation in clinical practice.

The authors also recognize a number of methodological limitations to this study.

- All the participants were stroke patients drawn randomly from the larger UKROC dataset, which collates a selected population of patients (mainly of working age) with complex neurological disabilities. These findings cannot necessarily be extrapolated to the more general population of stroke patients, who are mainly older with shorter lengths of stay in rehabilitation. Moreover, the present study focused solely on inpatients and it is possible that ceiling effects might be observed with a community sample post-discharge.
- A particular strength of our approach is that the analysis is based on the entire national dataset for stroke patients undergoing patient rehabilitation in Level 1 or 2 specialist rehabilitation units in England for the period, which means that the findings are likely to be generalizable for this population of patients. However, it should be noted that these are mainly tertiary rehabilitation services taking a selected group of (mainly younger) stroke patients with highly complex needs. Thus, further research on the UK FIM+FAM and the Rasch model with more diverse samples is indicated, as well as exploration in other patient groups (e.g. traumatic brain injury).
- The overall $\chi^2$ $p$-values for the final model were unable to detect errors if tested with random sample ($n=320$), but indicated errors if tested with the full sample. This suggests that the overall model errors are relatively small and may have appeared in the full sample due to methodological issues associated with RUMM2030, as suggested by thorough examinations (28, 29). This notion is also supported by satisfactory model fit reflected by other fit indices and all individual super-items in the final model tested with the full sample.

In conclusion, our analysis suggests that the UK FIM+FAM meets the Rasch model requirements with good reliability, acceptable targeting of each of the 3 domains, and with no item deletion in a population of complex stroke patients. A conversion table that accommodates DIF by stroke location has been produced, but this now requires further evaluation in clinical practice and in research.

## ACKNOWLEDGEMENTS

The authors would like thank all of the patients and clinicians who contributed to the UKROC dataset. Special thanks to Alan Tennant and Paula Kersten for their advice and guidance in the early stages of this analysis; to Heather Williams and Keith Sephton for their assistance with data extraction and cleaning; and to Roxana Vanderstay for her initial exploration of Rasch analysis in this programme. The authors would also like to thank Professor Svend Kreiner for his constructive review and advice on improving the manuscript.

This study was funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (RP-PG-0407-10185). Financial support for the preparation of this manuscript was also provided by the Dunhill Medical Trust.

*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, 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.
4. Hall KM, Mann N, High WMJ, et al. Functional measures after traumatic brain injury: ceiling effects of FIM, FIM+FAM, DRS, and CIQ. J Head Traum Rehabil 1996; 11: 27–39.
5. Turner-Stokes L, Nyein K, Turner-Stokes T, Gatehouse C. The UK FIM+FAM: development and evaluation. Clin Rehabil 1999; 13: 277–287.
6. Law J, Fielding B, Jackson D, Turner-Stokes L. The UK FIM+FAM Extended Activities of Daily Living module: evaluation of scoring accuracy and reliability. Disabil Rehabil 2009; 31: 825–830.
7. Specialist neuro-rehabilitation services: providing for patients with complex rehabilitation needs. London: British Society of Rehabilitation Medicine. Updated 2015. 2010. Available from: http://www.bsrm.org.uk/downloads/specialised-neurorehabilitation-service-standards--7-30-4-2015-forweb.pdf.
8. Turner-Stokes L, Williams H, Bill A, Bassett P, Sephton K. et al. 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.
9. Turner-Stokes L, Siegert RJ. A comprehensive psychometric evaluation of the UK FIM + FAM. Disabil Rehabil 2013; 35: 1885–1895.
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 Nation UK Clinical Dataset. PLOS One 2016; 11: e0147288.
11. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
12. Masters GA. Rasch model for partial credit scoring. Psychometrica 1982; 47: 149–174.
13. 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.
14. Bond TG, Fox JC. Applying the Rasch model. Fundamental measurement in the human sciences. 2nd edition. Lawrence Erlbaum Associates, Publishers: London; 2007.
15. Khan A, Chien CW, Brauer SG. Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. J Clin Epidemiol 2013; 66: 681–687.
16. Lundgren-Nilsson Å, Grimby G, Ring H, Tesio L, Lawton G, Slade A, et al. Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. J Rehabil Med 2005; 37: 23–31.
17. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatry 2010; 81: 1044–1048.
18. Lundgren-Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the Functional Independence Measure (FIM™) revisited. J Rehabil Med 2011; 43: 884–891.
19. Linn RT, Blair RS, Granger CV, Harper DW, O'Hara PA, Maciura E. Does the Functional Assessment Measure (FAM) extend the Functional Independence Measure (FIMTM) instrument? A Rasch analysis of stroke inpatients. J Outcome Measur 1999; 3: 339–359.
20. 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 Psychiatry 1999; 67: 749–754.
21. Mallinson T. Rasch Analysis of Repeated Measures. Rasch Measurement Transactions 2011; 251:1, 1317.
22. Lundgren-Nilsson A, Jonsdottir IH, Ahlborg G, Tennant A. Construct validity of the Psychological General Well Being Index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a Rasch analysis. Health Qual Life Outcomes 2013; 11: 2.
23. Marais I, Andrich D. Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. J Appl Meas 2008; 9: 105–124.
24. Andrich D, Hagquist C. Real and artificial differential item functioning in polytomous items. Educ Psychol Meas 2015; 75: 185–207.
25. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.
26. Leung Y-Y, Png M-E, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease – a special interest group report of OMERACT 11. J Rheumatol 2014; 41: 159–164.
27. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2030. Perth: RUMM Laboratory; 2010.
28. Müller M, Kreiner S. Item fit statistics in common software for Rasch analysis. Copenhagen, Denmark: Department of Biostatistics, University of Copenhagen; 2015 Jun. Available from: http://www.pubhealth.ku.dk/bs/publikationer.
29. 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.
30. Linacre JM. Sample size and item calibration stability. Rasch Meas Transact 1994; 7: 328.
31. Thurstone LL. The measurement of social attitudes. Abnormal Social Psychol 1931; 27: 249–269.
32. Turner-Stokes L, Fadyl J, Rose H. Williams H. Schulter P, McPherson KM. The work-ability support scale: evaluation of scoring accuracy and rater reliability. J Rehabil Med 2014; 24: 511–524.