J Rehabil Med 2005; 37: 346–352

DIFFERENTIAL ITEM FUNCTIONING OF THE FUNCTIONAL INDEPENDENCE MEASURE IN HIGHER PERFORMING NEUROLOGICAL PATIENTS

Annet J. Dallmeijer,[1,2] Joost Dekker,[1,2] Leo D. Roorda,[1,2] Dirk L. Knol,[2,3] Bianca van Baalen,[4] Vincent de Groot,[1,2] Vera P. M. Schepers[5] and Gustaaf J. Lankhorst[1,2] on behalf of the FuPro study group

From the [1]Department of Rehabilitation Medicine, [2]Institute for Research in Extramural Medicine (EMGO Institute) and [3]Department of Clinical Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, [4]Department of Rehabilitation Medicine, Erasmus Medical Center, Rotterdam, and [5]Department of Rehabilitation Medicine, University Medical Center, Utrecht, The Netherlands

Objective: **When comparing outcomes of the Functional Independence Measure (FIM™) between patient groups, item characteristics of the FIM™ should be consistent across groups. The purpose of this study was to compare item difficulty of the FIM™ in 3 patient groups with neurological disorders.**

Subjects: **Patients with stroke (*n* = 295), multiple sclerosis (*n* = 150), and traumatic brain injury (*n* = 88).**

Methods: **FIM™ scores were administered in each group. The FIM™ consists of a motor domain (13 items) and a cognitive domain (5 items). Rasch rating scale analysis was performed to investigate differences in item difficulty (differential item functioning) between groups.**

Results: **Answering categories of the FIM™ items were reduced to 3 (from the original 7) because of disordered thresholds and low answering frequencies. Two items of the motor domain ("bladder" and "bowel") did not fit the Rasch model. For 7 out of the 11 fitting motor items, item difficulties were different between groups (i.e. showed differential item functioning). All cognitive items fitted the Rasch model, and 4 out of 5 cognitive items showed differential item functioning.**

Conclusion: **Differential item functioning is present in several items of both the motor and cognitive domain of the FIM™. Adjustments for differential item functioning may be required when FIM™ data will be compared between groups or will be used in a pooled data analysis.**

Key words: disability evaluation, neurological disorders, Rasch analysis, rehabilitation.

J Rehabil Med 2005; 37: 346–352

Correspondence address: Annet J. Dallmeijer, Department of Rehabilitation Medicine, VU University Medical Center Amsterdam, PO Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: a.dallmeijer@vumc.nl

Submitted September 27, 2004; accepted April 1, 2005

INTRODUCTION

Measuring functional outcome in rehabilitation is important in both patient care and clinical research for evaluating the effectiveness of rehabilitation interventions and for predicting rehabilitation outcome. Several measurement instruments have been developed to measure functional outcome in neurological rehabilitation (1). Some of these instruments are generic measures that intend to measure the same construct (e.g. disability) across different patient groups. These measurement instruments usually consist of 1 or more subscales (domains), where items are summed to form a total subscale score.

When using the outcomes of generic measures in different patient groups, it is important that the items and subscales of these measurement instruments meet all (psychometric) assumptions regarding dimensional structure in each patient group. The dimensions should be consistent with the proposed dimensional structure of the measurement instrument, and items within dimensions should be measuring a single underlying construct. This is important when the scale intends to measure a single attribute or ability by adding item ratings to yield a total sum score (2). In this case, individual items of the same subscale are correlated with each other, and each item correlates with the total scale score it belongs to, and not to any other subscale (3). Measurement quality of the instrument can be investigated further by using Rasch analysis (4). This method, based on item response theory, converts ordinal scales into interval measures, allowing comparison among groups. It can be applied to determine differences in item difficulty within the subscales across patient groups, which is called differential item functioning (DIF) (5). The item difficulty should preferably be comparable across groups. If this is not the case, identical sum scores of 2 different patient groups are likely to result from different item profiles. As a consequence, they do not reflect the same level of functional status (e.g. level of independence) in different patient groups, which hampers comparison between groups (5).

The Functional Independence Measure (FIM™) is a generic measurement instrument that was developed in the USA to measure the severity of disability and is widely used to monitor progress during rehabilitation programs. The FIM™ intends to measure disability on 2 summated rating scales, a physical (motor) scale and a cognitive scale (6–9). The two-dimensional structure of the FIM™ has been confirmed in previous studies with heterogeneous groups of patients with various disorders (7, 10, 11). In addition, disease-specific studies have been performed in patients with stroke (12–14), traumatic brain injury

 ISSN 1650–1977
DOI 10.1080/16501970510038284

(TBI) (15) and multiple sclerosis (MS) (16). Other studies showed that the FIM^{TM} motor domain lacks unidimensionality (12, 13, 17). Granger et al. (6) and Heinemann et al. (10) investigated item difficulties of the FIM^{TM} across patient groups with Rasch analysis. They reported slight variations in item difficulties across 13 patient groups, but emphasized that these variations reflect clinical differences between patients. Also others reported dissimilarities in item difficulties of the FIM^{TM} and argued that caution should be taken when comparing results for different patient groups (18). However, until now, few studies have focused on detecting DIF across different patient groups, while FIM^{TM} data are often used in pooled analyses of various patient groups. In a recent publication (19), Rasch analysis was applied for cross-cultural validation of the motor domain of the FIM^{TM}, reporting DIF in 8 out of 13 motor items. These authors proposed a method for adjusting for DIF, allowing international comparisons and pooling of data of different patient groups.

The present study was performed to compare item difficulties (i.e. investigate DIF) of the FIM^{TM} in patients with stroke, MS and TBI, using Rasch analysis.

METHODS

Subjects

This project has been performed as part of a 3-year follow-up study on functional prognosis in 3 patient groups with different neurological disorders. The FIM^{TM} scores at 6 months after inclusion of the follow-up study were used for this study. The following patients were included in this study: (i) patients with a first-ever supratentorial stroke who were admitted for inpatient rehabilitation; (ii) patients with recently (<6 months) diagnosed MS; and (iii) patients with TBI admitted to the department of neurosurgery.

In total, 533 patients with neurological disorders were included. The group comprised 295 patients with stroke, 150 with MS and 88 with TBI. The characteristics of each patient group are listed in Table I.

Measures

The FIM^{TM} is a generic measurement instrument that records the degree of disability by evaluating the amount of assistance required to perform basic daily life activities (8, 9). It consists of 18 items, divided over 2 dimensions: the FIM-motor scale includes 13 items and the FIM-social-cognitive scale 5 items. Each item is measured on a 7-point rating scale, ranging from complete dependence (score 1) to complete independence (score 7). In this study, a Dutch translation of the FIM^{TM} was used.

FIM^{TM} item scores were collected by direct observation of and interviews with the patients or by interviewing proxies or caregivers. The FIM^{TM} was administered by trained clinical researchers (physiatrists).

Table I. *Subject characteristics for each patient group*

Characteristics	Stroke (n = 295)	MS (n = 150)	TBI (n = 88)
Age (years) (mean (SD))	57.5 (11.4)	38.3 (9.8)	35.3 (13.5)
Range (years)	18–80	18–65	17–67
Females (%)	40	63	27
Time since diagnosis at time of measurement (days) (mean (SD))	184 (25)	291 (62)	190 (28)
Range (days)	110–263	176–523	111–284

MS = multiple sclerosis; TBI = traumatic brain injury.

Statistics

Principal component analysis. Dimensionality refers to the assumption that items of a subscale measure the same underlying construct (20). Principal component analysis (PCA), followed by orthogonal (varimax) rotation, was performed to investigate the dimensionality of the FIM^{TM}. The number of factors to be rotated was restricted to the 2 proposed dimensions (7). Items were considered to load on a factor if factor loadings were higher than 0.40 on the proposed factor, and lower than 0.40 on the other factor (21). PCA was carried out using polychoric correlations instead of Pearson correlations, because the use of polychoric correlations is more appropriate in case of ordinal measures and skewed distributions (21). PCA was performed in each patient group separately.

Cronbach's α coefficients were calculated to determine internal consistency of the motor and cognitive subscales for each group separately. Scales were considered to be internally consistent when Cronbach's α was higher than 0.70 (20).

Rasch analysis. Rasch rating scale (22) analysis provides estimates of person ability and item difficulty along a common measurement continuum (4). Ordinal measures are transformed into linear continuous measures of person ability and item difficulty. Person ability and item difficulty are expressed in log-odd units (logits), which is a unit of interval measurement that is defined within the context of a set of items (5). Goodness of fit of an item set is determined by the fit statistics of the items. The fit statistic is an index for the consistency of the observed item score with the model expected scores: large misfit values indicate that the observed values of these items deviate from the model expected values based on the estimated person ability. High fit statistics (>1.7) (23) indicate that the observed item scores are much higher or lower than expected based on the item difficulties and estimated abilities of the subjects. Low statistics (<0.5) (23) indicates that items measure redundant or overlapping content area (5).

The analysis was performed on the pooled data first, to investigate whether all items fitted the Rasch model, and to identify possible disordered thresholds between answering categories, as recommended by Tennant et al. (19). If data fit the Rasch model, Rasch analysis allows detection of differences in item difficulties between groups, also referred to as differential item functioning (DIF). DIF was investigated by performing a Rasch analysis on each group separately, using items that fitted the model. To maintain comparability of the item difficulties of the subgroups, items were calibrated (anchored) using the step thresholds of the pooled data set (24). DIF was determined by comparing item difficulties between groups, using t-statistics. The motor and cognitive scales were examined separately. Mean Rasch measures (person abilities) for the motor and cognitive scales were calculated for each group by using the Rasch estimates for the whole group analysis (no correction for DIF was made). In addition, we calculated the Rasch measures (person abilities) of the motor and cognitive scales adjusted for DIF, by splitting the items that showed DIF in group specific items (i.e. an individual item is formed for stroke, MS or TBI), and than using these items for calculating the mean person abilities (19). Rasch analyses were performed using Bigsteps version 2.82 (24).

RESULTS

Subjects

Of the patients with stroke, 157 (53%) had left hemisphere lesions, 135 (46%) right hemisphere lesions and 3 (1%) had bilateral lesions. A total of 219 (74%) patients had a cerebral infarction and 76 (26%) patients had haemorrhagic strokes, of whom 30 (11%) had a subarachnoid haemorrhage. At the time of measurement, 32% of the patients were still receiving inpatient rehabilitation treatment. All other patients were living at home. Of the total of 150 MS patients, 115 (77%) had relapsing-remitting MS, 21 (14%) had primary-progressive MS at time of diagnosis, 8 (5%) had secondary-progressive MS, and in 6 (4%) patients the type of MS was unknown. All patients were living independently at home. Forty-one percent of the patients had received rehabilitation or paramedical treatment in the last month. The TBI group included 52 (59%) patients with a severe brain injury (Glasgow coma score (GCS) (25) at time of initial injury: 3–8), 17 (19%) patients with a moderate injury (GCS: 9–12), 12 (14%) patients with a mild injury (GCS: 13–15), and for 7 (8%) patients it was unknown. Forty-three percent of the TBI group had received inpatient rehabilitation treatment. At the time of measurement 9% of the patients received inpatient rehabilitation treatment, 8% stayed in a nursing home or hospital and 49% were living at home.

Functional Independence Measure

FIM-motor and cognitive raw scores of each patient group are shown in Table II. All groups had a high median score on both FIM-motor and FIM-cognitive subscales, indicating a high or moderate level of functional status. Ceiling effects larger than 15% were found in the motor scale for MS (23%) and TBI (26%), and in the cognitive scale for stroke (16%), MS (36%) and TBI (26%). Results of the PCA showed that all items of the FIM-motor scale loaded (i.e. were larger than 0.4) on the first factor (motor domain). The "eating" item of the MS group loaded on both factors, but the factor loading of the cognitive domain (0.41) was still lower than for the motor domain (0.54). The motor domain accounts for 47%, 39% and 54% of the total variance in stroke, MS and TBI, respectively. The FIM-cognitive items of the stroke, MS and TBI group all loaded on the second factor (cognitive domain), except for the "expression" item in MS that showed a factor loading just below 0.40 (0.39). The explained variance was 18%, 17% and 23% in stroke, MS and TBI, respectively.

Cronbach's α of the FIM-motor scale were 0.93, 0.89, and 0.98 for patients with stroke, MS, and TBI, respectively. For the FIM-cognitive scale Cronbach's α were 0.78, 0.68 and 0.88, respectively.

Rasch analysis: motor domain

The analysis of the pooled motor items revealed large misfit for 2 out of 13 motor items ("bladder" and "bowel"), showing fit statistics largely exceeding 1.7. In addition, disordered thresholds and low answering frequencies were found for the most dependent categories. When disordered thresholds and low answering categories are observed it is necessary to collapse adjacent categories (19). We therefore combined different answering categories. The most optimal solution (i.e. most items fitted the model) was tested by trial and error. We collapsed 5 adjacent answering categories (answering categories 1 through 5) to 1 category, resulting in a 3-category answering scale, ranging from dependence (score 1), modified independence (score 2) to complete independence (score 3). Results of the Rasch analysis of the collapsed (3-category) FIM-motor items for the total group showed that threshold estimates for the 3-category analysis were ordered. There was, however, still a considerable misfit of the "bladder" and "bowel" items (outfit statistics: 2.69 and 3.97). We therefore excluded these items from the further analysis, because items have to fit the Rasch model to investigate DIF (19). Following this, we performed the Rasch analysis with the 11 remaining items. All 11 items fitted the model (fit statistics <1.7), except for the "eating" item that showed a slight misfit (outfit statistics: 1.86, see Table III). After collapsing categories 1 through 5 a floor effect of 2.4% was found. The ceiling effect remained unchanged (see Table II).

Table II. *Raw FIM^{TM} motor and cognitive scores for each patient group*

	Stroke ($n=295$)	MS ($n=150$)	TBI ($n=88$)
FIM-motor			
Mean (SD)	76.2 (11.4)	86.7 (4.5)	83.0 (16.2)
Range	23–91	62–91	13–91
Median (inter quartile range)	79 (73–83)	88 (84–90)	89 (83–91)
Maximal score (%)	2	23	26
FIM-cognitive			
Mean (SD)	31.0 (3.8)	33.5 (1.6)	31.3 (4.9)
Range	17–35	28–35	9–35
Median (inter quartile range)	32 (29–34)	34 (33–35)	33 (30–35)
Maximal score (%)	16	36	26

MS = multiple sclerosis; TBI = traumatic brain injury.

To investigate DIF we performed the analysis for each group separately (Table III). To maintain comparability between groups, items were anchored using the threshold estimates of the total group. DIF plots of the motor items are shown in Fig. 1A. Rasch measures of each group are plotted against each other, with a line of identity and confidence interval. DIF was found for 7 out of the 11 motor items. The number of items showing DIF was the largest for stroke vs MS (7 items: 1, 2, 5, 6, 9, 12 and 13), but of the 7 items showing DIF only 4 showed considerable differences between groups (items: 1, 5, 9 and 12). The number of items showing DIF was 2 for MS vs TBI (items: 1 and 12), and 3 for stroke vs TBI (items: 1, 9 and 13). The motor items "bathing", "dressing upper body", "transfer tub" and "transfer toilet" showed no DIF.

Rasch analysis: cognitive domain

Results of the pooled Rasch analysis for the cognitive domain showed that all 5 items fitted the Rasch model. However, as in the motor domain, we found disordered thresholds and low answering frequencies in the most dependent categories. Because of this, and to maintain comparability with the motor domain we also combined the first 5 answering categories (answering categories 1 through 5) to 1 category in the cognitive domain. Results of the Rasch analysis of the collapsed (3-category) FIM-cognitive items for the total group are listed in Table IV. All five 3-category items fitted the Rasch model, and ordered threshold estimates were found.

DIF was investigated using the same procedure as described by the motor domain, performing Rasch analysis on each group separately (Table IV). DIF was found in 4 out of 5 items (Fig. 1B). The number of items showing DIF was 4 for stroke vs MS (item numbers: 1, 3, 4 and 5), 2 for stroke vs TBI (item

Table III. *Results of the Rasch analysis of the FIM™ motor scale (11 items)*[1]

	Total group		Stroke		MS		TBI	
FIM™ motor items	Item difficulty ±SE	Outfit	Item difficulty ±SE	Outfit	Item difficulty ±SE	Outfit	Item difficulty ±SE	Outfit
1 Eating	0.33±0.11	1.86	1.05±0.14	1.59	−1.84±0.31	1.60	−0.08±0.34	1.37
2 Grooming	−1.25±0.13	1.23	−1.11±0.15	1.02	−1.94±0.32	0.84	−0.99±0.38	1.51
3 Bathing	0.44±0.11	0.80	0.41±0.14	0.72	0.55±0.24	0.71	0.68±0.32	0.71
4 Dressing upper body	−0.22±0.12	0.88	−0.08±0.14	0.75	−0.37±0.26	1.03	−0.57±0.36	0.64
5 Dressing lower body	0.77±0.11	0.86	0.61±0.14	0.76	1.38±0.23	0.72	0.78±0.32	0.65
6 Toileting	−1.44±0.13	0.63	−1.22±0.15	0.64	−1.94±0.32	0.41	−2.19±0.45	0.74
9 Transfer bed, chair, wheelchair	−0.38±0.12	0.99	−0.86±0.15	0.63	0.72±0.24	0.72	0.15±0.33	1.51
10 Transfer toilet	−1.10±0.12	0.59	−1.13±0.15	0.47	−0.92±0.27	0.56	−1.29±0.40	0.79
11 Transfer tub, shower	−0.09±0.12	0.68	−0.12±0.14	0.66	−0.05±0.25	0.59	0.15±0.33	0.81
12 Walk	1.17±0.11	0.90	0.93±0.14	0.60	2.23±0.23	0.78	0.78±0.32	1.14
13 Stairs	1.76±0.11	0.84	1.51±0.14	0.73	2.18±0.23	0.62	2.57±0.31	1.32

[1] Item difficulty (and standard error, [SE]) and outfit statistics for the items of the FIM-motor scale are shown for each patient group. Item answering categories are reduced to 3 categories (see text).

numbers: 4 and 5), and 4 for MS vs TBI (item numbers: 1, 3, 4 and 5). The cognitive item “expression” showed no DIF.

Adjusted Rasch estimates

Raw FIM™ scores, Rasch estimates and adjusted Rasch estimates for the motor and cognitive domain, both expressed in original and standardized scores, are shown in Table V. For comparison reasons, standardized scores are calculated by transforming the raw scores and Rasch estimates to a scale ranging from 0 to 100. Pearson correlations between the adjusted and unadjusted Rasch measures exceeded 0.99.

DISCUSSION

Dimensional structure of the FIM™

Results of the principal component analysis confirmed the two-dimensional structure of the FIM™ in all patient groups. With a few exceptions, we found that all items of the FIM™ loaded on the expected factor. These findings are in agreement with earlier studies investigating the dimensional structure of the FIM™ in neurological patients (11). Sharrack et al. (16) investigated the FIM™ structure in 64 patients with MS. They also confirmed the two-dimensional structure, but reported that the cognitive domain accounted for only 6.4% of the variance and was not responsive. We also found a low factor loading for the “expression” item and a low Cronbach’s α (0.68) for the cognitive domain in MS, and to a lesser extent for stroke (0.68), which indicates the lower internal consistency of this scale. However, since the MS group had a ceiling effect of 36% in the cognitive dimension, these results should be further investigated in patients with lower levels of cognitive functioning.

In contrast to our results of the principal component analysis and the satisfactory Cronbach’s α, the more stringent Rasch analysis showed that goodness of fit was compromised by the “bladder” and “bowel” items. Previous studies applying Rasch analysis also reported considerable misfits for the “bladder” (7, 12–14, 18) and “bowel” (7, 13, 18) items in rehabilitation patients with different diseases combined (7) or patients with stroke separately (12–14, 18). This finding suggests that these 2 items measure a different construct than the other motor items. Kucukdeveci et al. (18) also mentioned that the misfit of the “bladder” and “bowel” items should be seen as “an inherent weakness of the FIM™”. However, we found a high correlation of 0.98 between total motor Rasch measures with and without the “bladder” and “bowel” items, indicating that excluding these misfitting items has only a minor effect on the total motor Rasch measure.

Item difficulty of the motor domain

Measurement quality of the FIM™ was also assessed through an analysis of DIF. Item difficulty is preferably similar for different patient groups (10, 26) because in that case equal sum scores reflect the same level of disability. We found that item difficulty varied across groups in 7 out of 11 motor items. For the motor domain, most items with DIF were found for the stroke-MS comparison, showing large DIF for 4 items, while comparison of the TBI group with stroke and MS revealed DIF in only 3 and 2 items, respectively. “Eating” was the only item showing DIF in all group comparisons. This finding should be interpreted with care, because the “eating” item was the only item that showed some misfit. Whether this is the consequence of the high functional level of our population (see below) should be further investigated. Patients with MS had more difficulty with “walking” than patients with stroke and TBI, and patients with stroke had less difficulty with the “transfer bed” item compared with patients from the other groups. In addition, patients with MS had more difficulty with “dressing lower body” than patients with stroke, and patients with TBI had more difficulty with “stairs” than those with stroke. Although these findings may reflect clinical differences between various neurological conditions (10), variations in item difficulty do limit comparability across groups. This is not in agreement with previous conclusions of

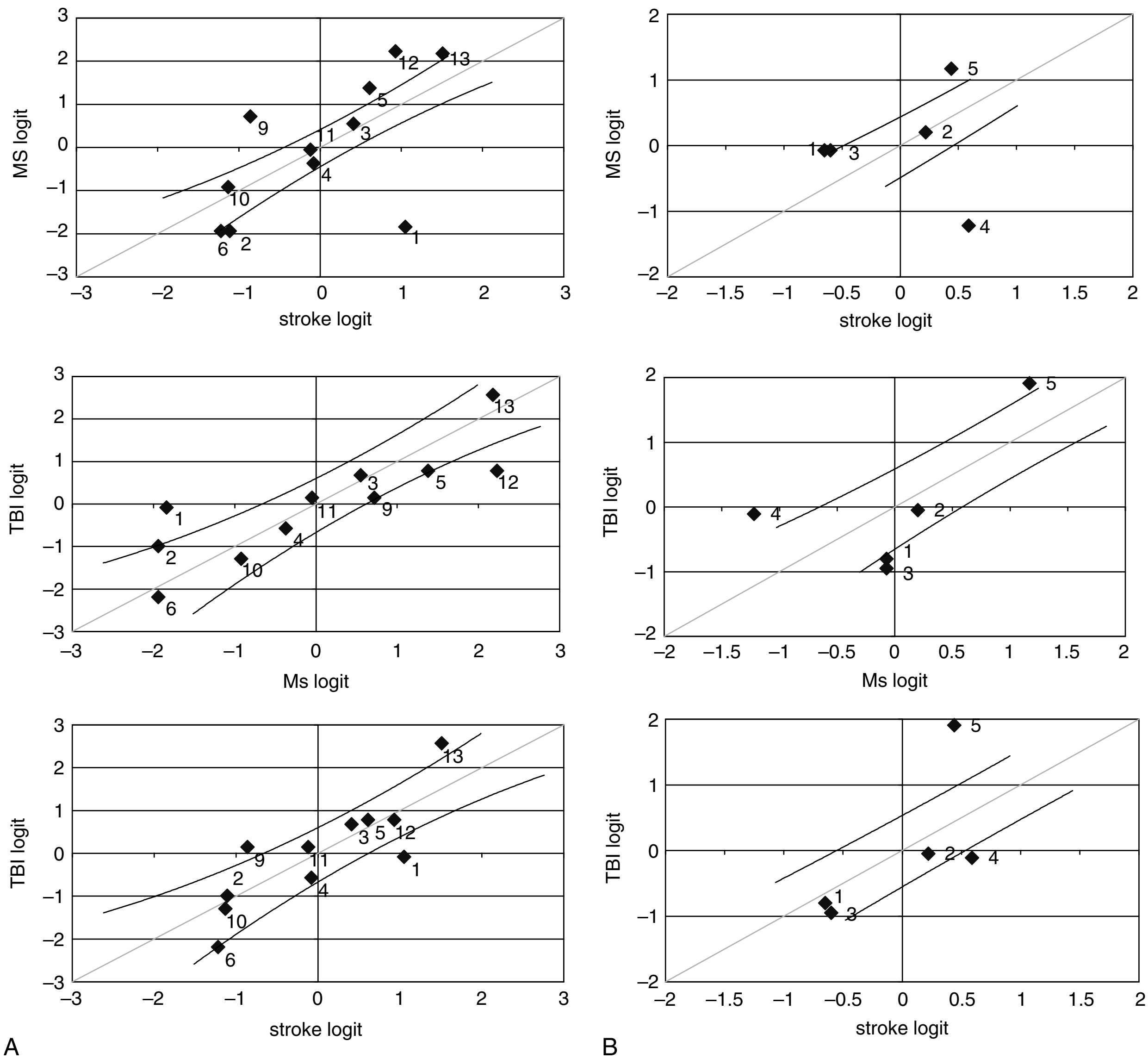

Fig. 1. Differential item functioning plots for the (A) motor and (B) cognitive FIM™ items for each patient group comparison. Item difficulties for 1 group are plotted on the x-axis and for the other group on the y-axis. An identity line is drawn through the origin with a slope of 1. The area between the 2 other lines indicate the 95% confidence interval. Items outside this area demonstrate DIF. MS = multiple sclerosis; TBI = traumatic brain injury.

Heinemann et al. (10) and Granger et al. (6), who compared item difficulties across patient groups in a large group of rehabilitation patients. They reported similar (but not identical) patterns for patients with different neurological disorders (stroke, brain dysfunction and neurological conditions). They did, however, not perform a DIF analysis.

Comparing item order of our study to results from the USA revealed some differences, particularly at the higher end of the scale (10). In the study Heinemann et al. reported that "eating", "bladder", "bowel" and "grooming" were the easiest motor items, and "stair climbing", "tub transfer" and "walking" the most difficult in all neurological groups. We found more or less the same item hierarchy at the lower end, but "eating", "bathing" and "dressing lower body" were among the most difficult items in our study, whereas "toilet transfer" was easier for our patients. However, the range between the easiest and most difficult items was in the present study smaller than usually reported (7, 10, 12), which may have influenced item hierarchy. Whether the differences between our findings and the results from the USA (10) are caused by cultural (or translation) differences (14, 18) should be addressed in future research.

Item difficulty of the cognitive domain

For the cognitive domain we found DIF for 4 out of 5 items, showing larger variations than in the motor domain. "Problem solving" and "memory" showed the largest differences in item

Table IV. *Results of the Rasch analysis of the FIM™ cognitive scale*[1]

FIM cognitive items	Total group		Stroke		MS		TBI	
	Item difficulty ± SE	Outfit	Item difficulty ± SE	Outfit	Item difficulty ± SE	Outfit	Item difficulty ± SE	Outfit
1 Comprehension	−0.56 ± 0.10	0.81	−0.65 ± 0.12	0.76	−0.07 ± 0.20	0.69	−0.80 ± 0.26	1.12
2 Expression	0.16 ± 0.10	0.86	0.22 ± 0.11	0.87	0.20 ± 0.19	0.92	−0.05 ± 0.24	0.83
3 Social interaction	−0.55 ± 0.10	1.15	−0.60 ± 0.12	1.20	−0.07 ± 0.20	0.92	−0.95 ± 0.27	1.13
4 Problem solving	0.14 ± 0.10	0.88	0.59 ± 0.11	0.84	−1.22 ± 0.26	0.87	−0.11 ± 0.25	0.88
5 Memory	0.80 ± 0.10	1.32	0.44 ± 0.11	1.48	1.17 ± 0.17	0.60	1.91 ± 0.24	1.02

[1] Item difficulty (and standard error [SE]) and outfit statistics for the items of the FIM cognitive scale are shown for each patient group. Item answering categories are reduced to 3 (see text).

Table V. *Standardized FIM™ raw scores, standardized*[1] *Rasch estimates and standardized Rasch estimates adjusted for differential item functioning for the motor and cognitive domain in each patient group*[2]

	Stroke ($n = 295$)		MS ($n = 150$)		TBI ($n = 88$)		Total group ($n = 533$)	
	Mean (SD)	Standardized mean (SD)	Mean (SD)	Standardized mean (SD)	Mean (SD)	Standardized mean (SD)	Mean (SD)	Standardized (mean (SD))
FIM-motor (11 items)								
Raw score	22.0 (5.4)	50.1 (24.7)	29.7 (3.5)	84.9 (15.9)	28.4 (6.0)	79.1 (27.3)	25.2 (6.2)	64.7 (28.2)
Rasch estimate	−0.02 (2.90)	49.8 (23.3)	4.16 (1.98)	83.3 (15.9)	3.47 (3.32)	77.8 (26.7)	1.73 (3.38)	63.8 (27.1)
Adjusted Rasch estimate	−0.06 (2.96)	48.9 (22.4)	4.38 (2.22)	82.5 (16.8)	3.57 (3.47)	76.3 (26.2)	1.78 (3.54)	62.9 (26.7)
FIM-cognitive (5 items)								
Raw score	11.5 (2.8)	65.4 (27.6)	13.5 (1.6)	84.8 (16.3)	12.1 (3.0)	71.1 (29.5)	12.2 (2.7)	71.8 (26.6)
Rasch estimate	0.97 (1.74)	64.4 (25.9)	2.20 (1.16)	82.6 (17.2)	1.36 (1.91)	70.1 (28.4)	1.38 (1.71)	70.5 (25.4)
Adjusted Rasch estimate	0.99 (1.80)	59.2 (23.8)	2.27 (1.23)	76.1 (16.3)	1.75 (2.2)	69.3 (28.9)	1.47 (1.83)	65.6 (24.1)

[1] Rasch estimates (person abilities expressed in logits) are transformed to a 0–100 scale; [2] Item answering categories are reduced to 3.

difficulty, while "expression" was the only item without DIF. "Problem solving" was the most difficult item in patients with stroke, whereas it was the easiest item for patients with MS. "Memory" was on the other hand easier for patients with stroke than for patients with MS and TBI. Heinemann et al. (10) also distinguished 3 item difficulty patterns for the FIM-cognitive domain in neurological patient groups. These findings suggest that comparability of the FIM-cognitive domain between these patient groups is limited and can only be performed when items are adjusted for DIF.

Adjusted Rasch estimates

Our results indicate that when comparing or pooling data of patients with neurological disorders, in particularly for patients with stroke and MS, adjustments for DIF may be necessary. This accounts to a lesser extent for comparisons of TBI with stroke and MS. Tennant et al. (19) recently described a method for such adjustments, in which the group-specific Rasch measures are used in the analysis for items showing DIF. In this procedure, for all items showing DIF, group specific items are used in the calculation of the total Rasch estimates. In our study, for example, we split item 9 ("transfer bed") into 3 group-specific items for stroke, MS and TBI. We found some differences between adjusted and unadjusted Rasch measures (Table V), but Pearson correlation between the adjusted and unadjusted Rasch measures were high (exceeding 0.99). This indicates that adjusting for DIF seems to have only minor impact on the person abilities in this cross-sectional design. It is, however, important to further explore the benefits of using adjusted Rasch measures for measuring improvements in rehabilitation using the FIM™. To what extent the corrected Rasch measures may improve the responsiveness of the FIM™ should for example be investigated in future studies.

Limitations of the study

The relatively good functional status of the investigated patient groups has resulted in a skewed data distribution towards the higher end of the scale (see Table II). This limits generalization of the results to persons with lower levels of functional status. Nevertheless, it has been argued that the FIM™ should also be valid for subjects with higher functional abilities (6). It is acknowledged that, at the lower and higher end of the scoring range, Rasch estimates are more accurate indicators of person ability (or change in ability) than the raw FIM™ scores (6, 12). It may therefore be recommended to use Rasch estimates of person ability instead of raw FIM™ scores to analyse results of groups with high functional abilities.

In addition, for performing Rasch analysis in the present study, it was necessary to collapse the lower answering categories due to low answering frequencies and disordered thresholds. The low frequency of the most dependent answering categories is also the consequence of the high functional level of the investigated groups. Although the collapsing of the answering categories reduces comparability with other FIM™ data, it may be an appropriate procedure in better functioning patients. Nevertheless, disordered thresholds of the 7 FIM™ categories have been reported before in a large FIM™ study (19), indicating that this finding is not a peculiarity of the present study.

Because of the small number of subjects in the TBI group, the results for this group should be interpreted with caution. Despite this limitation, the results indicate that for the motor domain only slight DIF is present for comparison of TBI with both stroke and MS. However, the large confidence interval due to the small number of subjects should be taken into account, because this reduces the number of items identified as having DIF. In the cognitive domain, DIF was identified in several items for comparison with both stroke and MS. These findings should be confirmed in future studies with a larger sample size.

CONCLUSION

In conclusion, the purpose of this study was to investigate whether the FIM™ motor and cognitive domains can be used for comparing disability in patients with different neurological disorders. The "bladder" and "bowel" items of the motor domain showed lack of fit to the model, and were therefore excluded from the DIF analysis. Item difficulties of the motor domain showed DIF in 7 items, being most apparent in the comparison of patients with stroke and MS. All items of the cognitive domain fitted the Rasch model, but there were large differences in item difficulty for the "problem solving" and "memory" items. It is concluded that adjustments for DIF may be required when comparing or pooling data of the FIM™ motor and cognitive domains in patients with stroke, MS and TBI.

ACKNOWLEDGEMENTS

This project has been performed as part of the "Functional prognostication and disability study on neurological disorders", supervised by the department of Rehabilitation Medicine of the VU Medical Center, Amsterdam and was funded by the Netherlands Organization for Health Research and Development (grant: 96-06-002).

FuPro study group: G. J. Lankhorst, J. Dekker, A. J. Dallmeijer, M. J. IJzerman, H. Beckerman, V. de Groot: VU University Medical Center Amsterdam (project coordination); A. J. H. Prevo, E. Lindeman, V. P. M. Schepers: University Medical Center, Utrecht; H. J. Stam, E. Odding, B. van Baalen: Erasmus Medical Center, Rotterdam; A. Beelen: Academic Medical Center, Amsterdam, The Netherlands.

REFERENCES

1. Wade DT. Measurement in neurological rehabilitation. Oxford: Oxford University Press; 1992.
2. Silverstein B, Kilgore KM, Fisher WP, Harley JP, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: I. Exploring unidimensionality. Arch Phys Med Rehabil 1991; 72: 631–637.
3. Streiner DL, Norman GR. Health measurement scales. 2nd edition. Oxford: Oxford Medical Publications; 1995.
4. Wright BD, Linacre JM, Smith RM, Heinemann AW, Granger CV. FIM measurement properties and Rasch model details. Scand J Rehabil Med 1997; 29: 267–272.
5. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35: 105–115.
6. Granger CV, Hamilton BB, Linacre JM, Heinemann AW, Wright BD. Performance profiles of the functional independence measure. Am J Phys Med Rehabil 1993; 72: 84–89.
7. 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.
8. Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Teshman JS. Rehabilitation outcomes: analysis and measurements. In: Fuhrer MJ, ed. A uniform national data system for medical rehabilitation. Baltimore: Brookes; 2003, pp. 137–147.
9. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. Adv Clin Rehabil 1987; 1: 6–18.
10. Heinemann AW, Linacre JM, Wright BD, Hamilton BB, Granger C. Relationships between impairment and physical disability as measured by the functional independence measure. Arch Phys Med Rehabil 1993; 74: 566–573.
11. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil 1996; 77: 1101–1108.
12. Brock KA, Goldie PA, Greenwood KM. Evaluating the effectiveness of stroke rehabilitation: choosing a discriminative measure. Arch Phys Med Rehabil 2002; 83: 92–99.
13. Chang WC, Chan C. Rasch analysis for outcomes measures: some methodological considerations. Arch Phys Med Rehabil 1995; 76: 934–939.
14. Tsuji T, Sonoda S, Domen K, Saitoh E, Liu M, Chino N. ADL structure for stroke patients in Japan based on the functional independence measure. Am J Phys Med Rehabil 1995; 74: 432–438.
15. 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.
16. Sharrack B, Hughes RA, Soudain S, Dunn G. The psychometric properties of clinical rating scales used in multiple sclerosis. Brain 1999; 122: 141–159.
17. Dickson HG, Kohler F. The multi-dimensionality of the FIM motor items precludes an interval scaling using Rasch analysis. Scand J Rehabil Med 1996; 28: 159–162.
18. Kucukdeveci AA, Yavuzer G, Elhan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. Clin Rehabil 2001; 15: 311–319.
19. Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004; 42 (suppl): I37–I48.
20. Nunnally JC, Bernstein IH. Psychometric theory. 3rd edition. New York: McGraw-Hill; 1994.
21. Kim JO, Mueller ChW. Introduction to factor analysis what it is and how to do it. Beverly Hills, CA: Sage Publications; 1996.
22. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978; 43: 561–573.
23. Wright BD, Linacre JM, Gustafson JE, Martin-Löf P. Reasonable mean-square fit values. Rasch Measurement Transactions 1994; 8 (3): 370.
24. Linacre JM, Wright BD. A user's guide to BIGSTEPS: Rasch-model computer program. Chicago: MESA Press; 1998.
25. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974; 2: 81–84.
26. Silverstein B, Fisher WP, Kilgore KM, Harley JP, Harvey RF. Applying psychometric criteria to functional assessment in medical rehabilitation: II. Defining interval measures. Arch Phys Med Rehabil 1992; 73: 507–518.