

## Original Article

# Cross-cultural validity of FIM in spinal cord injury

G Lawton<sup>1</sup>, Å Lundgren-Nilsson<sup>2</sup>, F Biering-Sørensen<sup>3</sup>, L Tesio<sup>4</sup>, A Slade<sup>5</sup>, M Penta<sup>6</sup>, G Grimby<sup>2</sup>, H Ring<sup>7</sup> and A Tennant<sup>\*1</sup>

<sup>1</sup>Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds, Leeds, UK; <sup>2</sup>Department of Rehabilitation Medicine, Sahlgrenska Academy at Göteborg University, Sweden; <sup>3</sup>Clinic for Para- and Tetraplegia, The Neuroscience Centre, Rigshospitalet, Copenhagen University Hospital, Denmark; <sup>4</sup>Unit of Rehabilitation Research, Istituto Auxologico Italiano, Milan, Italy; <sup>5</sup>School of Health and Human Sciences Leeds Metropolitan University, Leeds, UK; <sup>6</sup>Unité de Réadaptation et de Médecine Physique, Université Catholique de Louvain, Bruxelles, Belgium; <sup>7</sup>Loewenstein Hospital, Rehabilitation Centre, Raanana, Tel Aviv University, School of Medicine, Israel

**Objective:** To analyse cross-culture validity of the Functional Independence Measure (FIM™) in patients with a spinal cord injury using a modern psychometric approach.

**Settings:** A total of 19 rehabilitation facilities from four countries in Europe.

**Participants:** A total of 647 patients at admission, median age 46 years, 69% male.

**Methods:** Data from the FIM™, collected on inpatient admission, was fitted to the Rasch model. A detailed analysis of scoring functions of the seven categories of the FIM™ items was undertaken before to testing fit to the model. Categories were rescored where necessary. Fit to the model was assessed initially within country, and then in the pooled data. Analysis of differential item functioning (DIF) was undertaken in the pooled data for each of the FIM™ motor and social cognitive scales, respectively. Final fit to the model was tested for breach of local independence by principle components analysis (PCA).

**Results:** The present scoring system for the FIM™ motor and cognitive scales, that is a seven category scale, was found to be invalid, necessitating extensive rescoring. Following this, DIF was found in a number of items within the motor scale, requiring a complex solution of splitting items by country to allow for the valid pooling of data. Five country-specific items could not be retained within this solution. The FIM cognitive scale fitted the Rasch model after rescoring, but there was a substantial ceiling effect.

**Conclusions:** Data from the FIM™ motor scale for patients with spinal cord injury should not be pooled in its raw form, or compared from country to country. Only after fit to the Rasch model and necessary adjustments could such a comparison be made, but with a loss of clinical important items. The FIM cognitive scale works well following rescoring, and data may be pooled, but many patients were at the maximum score.

*Spinal Cord* (2006) 44, 746–752. doi:10.1038/sj.sc.3101895; published online 3 January 2006

**Keywords:** spinal cord injury; assessment; rehabilitation; rasch analysis

## Introduction

The use of generic outcome measures is seen as a common part of the management of spinal cord injury (SCI).<sup>1</sup> Although many such measures are available, their use in clinical practice throughout Europe is limited.<sup>2</sup> While the demands of clinical management in a hospital setting may indicate a preference for a given instrument, there are several factors that may determine which, if any generic measure is chosen. For example, within Europe, outcome measures will need to be

adapted to a particular language,<sup>3</sup> and thus, preference may be given to those outcome measures that already have a local adaptation.

If data are to be pooled to facilitate European-based comparisons of health care, a requirement is for outcome measures to work in a consistent manner across countries to demonstrate cross-cultural validity. A project called 'European Standardisation of Outcome Measurement in Rehabilitation' (Pro-ESOR) was established under the Framework IV programme of the European Commission (EC) to examine the internal- and cross-cultural validity of commonly used outcome measures. In patients with SCI, Manual Muscle Testing,

\*Correspondence: A Tennant, Academic Unit of Musculoskeletal & Rehabilitation Medicine, The University of Leeds, 36 Clarendon Road, Leeds LS2 9NZ, UK

the Functional Independence Measure, and the Ashworth Scale for Spasticity (different modifications) were the most widely used assessments.<sup>2</sup> Using the terminology from the International Classification of Functioning, Disability and Health (ICF)<sup>4</sup> the first of these is a physiological assessment of impairment; the latter two are outcome scales for activity limitation and impairment, respectively.

This paper is concerned with the cross-cultural validity of the FIM™ in SCI. The FIM™ is a measure of activity limitation that is used across a wide range of conditions and in a variety of situations in rehabilitation. There is an extensive body of literature supporting reliability, validity, and the responsiveness of FIM™, although the latter may vary with the population being assessed.<sup>5</sup> Through the original recommendation of the American Spinal Cord Injuries Association (ASIA) the FIM™ has seen widespread use in outcomes research.<sup>6–9</sup> Segal *et al*<sup>10</sup> found high reliability on total FIM™ scores between two spinal cord injury rehabilitation facilities, though this varied in terms of individual item scores. In particular, the social cognitive items showed low reliability, which the authors suggested may be related to the complexity and lack of understanding of these items. Others<sup>11</sup> have demonstrated that the social cognitive items cannot be used as a substitute for a comprehensive neuropsychological assessment, and the high percentage of ceiling scores found on these items in a number of studies<sup>12</sup> suggests that they are not sensitive enough for this patient group.

The focus of this paper is purely on the cross-cultural validity of the FIM™ by analysing pooled data from a Rasch measurement model perspective. It will be demonstrated that pooling of raw score data from the motor scale across countries is not valid but, after necessary adjustments, comparison between different countries can be made. The validity of the cognitive scale for this diagnostic group is questioned.

## Methods

### *Patients and settings*

Nineteen rehabilitation facilities within four different countries contributed anonymous data from patients recently admitted with a diagnosis of SCI. The only data required from the facilities were raw scores on each item of the FIM™ scale at admission, together with the age and gender of the patient.

### *Functional independence measure*

The FIM™ consists of 13 motor and five social cognitive items, assessing self-care, sphincter management, transfer, locomotion, communication, social interaction, and cognition.<sup>6</sup> It uses a 7-level scale anchored by extreme rating of total dependence as 1 and complete independence as 7; the intermediate levels are as follows: 6 modified independence, 5 supervision or setup, 4 minimal contact assistance or the subject expends >75% of the effort, 3 moderate assistance or the subjects expends

50–75% of the effort, and 2 maximal assistance or the subject expends 25–50% of the effort. Although developed originally as an 18 item scale, it has been shown that there are two scales, a 13 item motor and a 5 item social-cognitive scale.<sup>13</sup> In the present study, the original scales will be referred to as FIM™ motor and FIM™ social cognitive items or scales, respectively. Owing to copyright issues, if once changed in any way, these will be referred to as the FIM motor and FIM social cognitive scales.

### *Rasch analysis*

The Rasch model is used as a methodological basis for examining the internal construct validity of a scale, its scaling properties, and cross-cultural validity through fitting data from the scale to the Rasch model. It is a unidimensional measurement model, which assumes that the easier the item the more likely that it will be passed, and the more able the person, the more likely that they will pass an item compared to a less able person.<sup>14</sup> In other words, there is the probability, in the dichotomous case, that a person will affirm an item is a logistic function of the difference between the person's ability ( $\theta$ ) and the difficulty of the item ( $b$ ), and only a function of that difference. From this, the expected pattern of responses to an item set is determined given the estimated  $\theta$  and  $b$ . When the observed response pattern coincides with or does not deviate too much from the expected response pattern, then the items constitute a unidimensional measure. Taken with confirmation of local independence of items, that is, no residual associations in the data after the Rasch trait has been removed, this confirms unidimensionality.<sup>15</sup> For cases where items have more than two categories, the model includes an explicit 'threshold' parameter ( $\tau$ ),<sup>16</sup> where the threshold represents the equal probability point between any two adjacent categories within an item. In the logit form

$$\ln\left(\frac{P_{nik}}{1 - P_{nik-1}}\right) = \theta_n - b_i - \tau_k$$

giving the log-odds of person  $n$  affirming category  $k$  in item  $i$ ,  $\theta$  is person ability,  $b$  is the item difficulty parameter,  $\tau_k$  is the difficulty of the  $k$  threshold, and  $P_{nik}$  is the probability for person  $n$  to answer item  $i$  in category  $k$ . The units of measurement obtained from the equation are called 'logits' which is a contraction of log-odds probability units. These threshold estimates should be correctly ordered if the categories are being assigned in the intended way. Consequently, this can be empirically verified against the model expectation and deviations identified where the categories fail to express an increasing level of the trait (disordered thresholds).

Where disordered thresholds are identified, the discrimination of each category is examined. Choices of the way in which categories are collapsed are based on the category discrimination, or where unclear, on clinical knowledge. Once disordered thresholds are removed, fit of data to the Rasch model is assessed in

a number of ways. Item fit statistics indicate how well the items are fitting the model individually. These are given in the form of residuals value (the standardised difference between the observed and the expected score for each person), which should be between  $-2.5$  and  $2.5$ , and  $\chi^2$  statistics, which should show nonsignificant deviation from the model. The  $\chi^2$  values are calculated based on ability groups (or class intervals) of approximately 50 people, which the patients are assigned to based on their total score. Overall fit of the scale as a whole is also given by standardised fit statistics for persons and items (mean zero, SD of one where the data fit the model perfectly), and a  $\chi^2$  Item-Trait interaction statistic (calculated by summing all the item  $\chi^2$  values and degrees of freedom) to determine scale invariance that once again should indicate nonsignificant deviation from the model.

A principle components analysis (PCA) is performed on the person-item residuals to confirm local independence, that is, no pattern in the residuals following the removal of the Rasch factor.

Much of the published work on Rasch analysis in rehabilitation has explored issues of unidimensionality and scaling properties.<sup>17–19</sup> However, Rasch analysis allows for much more than an empirical test for unidimensionality. The basis of the approach to the analysis of cross-cultural validity lies in the plot of the proportion of individuals *at the same ability level* (grouped into class intervals) who answer a given item correctly (or can perform a particular task). These proportions, except for random variations, should be the same irrespective of the nature of the group for whom the proportions are plotted.<sup>20</sup> Items that do not yield the same plot for two or more groups display differential item functioning (DIF) and are violating the requirement of unidimensionality. Consequently it is possible to examine whether or not a scale works in the same way by contrasting the response function for each item across cultures. This process has been described in more detail in another paper from our group, including the procedure for rescoring the responses.<sup>21</sup> DIF may manifest itself as a constant difference between countries across the trait (Uniform DIF – the main effect), or as a variable difference, where the response function of the two groups cross over (nonuniform DIF – the interaction effect). Both the country factor and the interaction with the class interval might be significant in some cases, as with any ANOVA's main and interaction

effects. Tukey's *post hoc* tests determine where the statistically significant differences are to be found where there are more than two groups.

Where some but not all items display DIF, it is possible to make an adjustment to allow items with DIF to vary by country. To do this, an item is substituted for a series of country-specific items (eg Bathing becomes Bathing – Israel, Bathing – Italy, etc.). For each country, only the scores observed in its corresponding item are considered, while the other items are assigned missing values. Subsequent analysis is undertaken on this expanded data set (ie original plus split items). Finally, the tests for local independence are undertaken to confirm unidimensionality of the scale.<sup>22</sup>

Many of the fit statistics for Rasch analysis are  $\chi^2$  based, even those based on the likelihood of the data capitalizing on the fact that a  $-2 \log$  (likelihood) is asymptotically  $\chi^2$ .<sup>23</sup> Given that tests of fit are set against 'perfect fit', the evidence appears to be that sample sizes between 50 and 250 data points are appropriate. However, if many patients are at the margins, and thus, the scale is poorly targeted, this also has an influence on sample size.<sup>23</sup> Consequently, if a scale is well targeted (ie 40–60% success rates on dichotomous test items), then a sample size of 108 will give 99% confidence of being within  $\pm 0.5$  logits. If not well targeted (ie  $< 15\%$  or  $> 85\%$  success rate), this rises to 243. For tests for differential item functioning, a sample size of 200 or less has been suggested as adequate.<sup>24</sup> For example, 150 cases per country are sufficient to test for DIF where at  $\alpha$  of 0.01 a difference of 0.5 logits within the residuals can be detected for any two groups with  $\beta$  of 0.20.<sup>25</sup>

The Rasch analysis was undertaken with the RUMM2020 software.<sup>26</sup> Owing to the number of tests of fit undertaken (eg 13 for each item in the motor scale), Bonferroni corrections were applied, giving a significant *P*-value of 0.004 for the motor FIM<sup>TM</sup> and 0.01 for the social cognitive FIM<sup>TM</sup>.<sup>27</sup>

## Results

### Participants

A total of 647 patients were recruited with a mean age, which varied between 35 and 57 years (range 11–93) across the countries, of whom 31% were female (Table 1). For pooled data, the common median age of 46 years is used for the whole sample. The number of

**Table 1** Characteristics of patients recruited into the study

Country	Number of patients	% Females	Number of contributing institutions	Mean age	Median age	Range	Mean FIM score	Range
Denmark	168	33	1	49	49	13–93	50.6	13–91
Israel	153	25	4	39	35	11–75	39.2	13–90
Italy	226	36	10	51	57	14–91	38.2	13–91
UK	100	25	4	43	43	15–81	50.1	13–89
Total	647	31	19	46	46	11–93	43.1	13–91

contributing institutions varied from Italy having 10 contributing institutions to Denmark with only one contributing institution. The mean admission FIM™ Motor Score was 43.1, with a significant difference between countries ( $F=9.732$ ;  $P<0.001$ ). *Post hoc* tests showed that Israel and Italy had much lower admission Motor scores (about 10 points) than the UK or Denmark. The mean admission FIM™ Cognitive Score was 32.3, again with a significant difference across countries ( $F=6.981$ ;  $P<0.001$ ), but on this occasion the difference was between the UK and some other country.

#### *Pooled data and cross-cultural validity*

When data from the FIM™ motor scale from the four countries were pooled, 12 of the items displayed disordered thresholds and needed to be rescored. Following this, fit to the model was poor, with seven items showing significant misfit (Table 2). An examination of DIF found little evidence for age or gender for the motor scale, with just two items. Transfer bed and Walk/Wheelchair showing nonuniform DIF by age and gender, respectively. In contrast, eight items showed DIF by country, and Tukey's *post hoc* comparison of these items showed a complex pattern of item-country DIF (Table 3). Consequently, where an item worked differently in different countries, it was split into four

separate items, one for each country. This resulted in 37 items (five original items and eight items split for each country). These data were refitted to the Rasch model. The item 'Transfer tub/shower UK' was removed from the analysis, as it was an extreme item (ie, all responses were at the extreme ends of the response categories). After rescoring for disordered thresholds, five misfitting items needed to be removed from the scale; these were 'Dressing upper body-UK', 'Bladder management Denmark', 'Walk/wheelchair Denmark', 'Eating' and 'Grooming' (Table 4). Following this, the items showed good fit at the individual level, though overall item trait interaction still showed significant deviation from model expectations ( $\chi^2=60.056$ ,  $df=31$ ,  $P=0.0014$ ). PCA of the residuals indicated that the first factor explained 17% of the variation, and the second factor 15%, indicating little substantive patterning in the residuals and thus supporting the unidimensionality of the scale.

The FIM cognitive scale was found to fit the model well after rescoring items, both at the individual item level (Table 5) and overall ( $\chi^2=5.121$ ,  $df=5$ ,  $P=0.390527$ ). No DIF by age, gender or country was observed. The PCA analysis of the residuals showed a first factor accounting for 34% of the variation, and the second 28%, again supporting the unidimensionality of the original Rasch factor.

**Table 2** Individual motor item fit across all countries using pooled data after rescoring

Item	Number of Categories	Item difficulty (logits)	SE	Residual	$\chi^2$	Probability
Eating	2	-6.16	0.25	-0.35	3.84	0.2799
Grooming	2	-5.07	0.19	-0.89	3.44	0.3284
Bathing	7	-0.60	0.05	-0.68	6.21	0.1019
Dressing upper body	4	-1.47	0.07	-0.12	10.54	0.0145
Dressing lower body	2	2.41	0.16	-1.27	16.04	<b>0.0011</b>
Toileting	2	1.56	0.14	-2.00	24.96	<b>&lt;0.004</b>
Bladder management	2	1.37	0.14	1.05	23.12	<b>&lt;0.004</b>
Bowel management	2	1.51	0.14	<b>2.95</b>	8.72	0.0333
Transfer bed	4	0.01	0.07	-2.13	19.23	<b>0.0002</b>
Toilet	2	2.25	0.15	-1.58	21.18	<b>0.0001</b>
Transfer tub/shower	4	0.77	0.08	-2.02	5.60	0.1326
Walk/wheelchair	3	0.30	0.11	<b>3.18</b>	11.00	0.0117
Stairs	3	3.13	0.13	-0.15	2.32	0.5085

Residuals and probabilities in bold indicates misfit to the model. Bonferroni corrected  $P=0.004$

**Table 3** Tukey's *post hoc* comparison for seven motor items with DIF by country

Comparing countries	Bathing	Dressing upper body	Dressing lower body	Bladder management	Transfer toilet	Transfer tub	Walk
Denmark-Israel	*	*	*	*	*		
Denmark-Italy				*	*	*	*
Denmark-UK		*	*	*			
Israel-Italy	*	*				*	*
Israel-UK	*						
Italy-UK							*

\*Indicates a significant difference

**Table 4** Fit of FIM Motor scale following adjustment for DIF

<i>Item</i>	<i>Number of Categories</i>	<i>Item difficulty (logits)</i>	<i>SE</i>	<i>Residual</i>	$\chi^2$	<i>Probability</i>
Bathing Denmark	7	-2.21	0.11	0.21	1.37	0.2416
Bathing Israel	7	-2.95	0.11	0.86	4.54	0.0331
Bathing Italy	7	-1.81	0.09	0.20	0.27	0.6021
Bathing UK	5	-0.91	0.16	0.57	0.04	0.8348
Dressing upper body Denmark	4	-4.45	0.15	-0.69	0.46	0.4966
Dressing upper body Israel	3	-2.03	0.18	-0.27	0.61	0.4337
Dressing upper body Italy	7	-2.81	0.09	-0.65	2.45	0.1175
Dressing lower body Denmark	2	0.22	0.28	-1.32	1.74	0.1878
Dressing lower body Israel	2	2.12	0.45	-0.51	1.08	0.2996
Dressing lower body Italy	2	1.60	0.35	-1.02	1.56	0.2118
Dressing Lower Body UK	2	2.02	0.45	0.13	4.04	0.0444
Toileting	2	0.35	0.15	-2.10	7.29	0.0069
Bladder management Israel	2	-0.43	0.27	0.18	0.37	0.5453
Bladder management Italy	2	-0.25	0.25	-0.01	0.11	0.7376
Bladder management UK	2	-0.38	0.32	0.82	1.58	0.2083
Bowel management Denmark	2	1.55	0.30	0.57	10.31	0.0013
Bowel management Israel	2	0.11	0.29	0.29	0.23	0.6337
Bowel management Italy	2	-0.56	0.24	1.32	0.01	0.9210
Bowel management UK	2	0.18	0.33	0.13	0.20	0.6515
Transfer bed	4	-1.35	0.08	-2.30	6.71	0.0096
Transfer toilet Denmark	2	0.13	0.28	-1.57	1.82	0.1772
Transfer toilet Israel	2	1.48	0.39	-0.57	0.76	0.3833
Transfer toilet Italy	2	2.12	0.40	-0.83	0.60	0.4393
Transfer toilet UK	2	1.03	0.36	-1.15	0.61	0.4360
Transfer tub/shower Denmark	4	-1.23	0.16	-1.79	2.07	0.1503
Transfer tub/shower Israel	4	-0.99	0.16	-1.75	0.90	0.3432
Transfer tub/shower Italy	4	0.51	0.16	-0.70	0.17	0.6808
Walk/wheelchair Israel	2	1.05	0.35	-0.12	0.94	0.3313
Walk/wheelchair Italy	2	3.21	0.55	0.61	6.23	0.0126
Walk/wheelchair UK	2	2.60	0.52	-0.65	0.13	0.7240
Stairs	3	2.08	0.13	0.73	0.86	0.3548

Bonferroni corrected  $P = 0.004$ **Table 5** Fit of FIM cognitive scale

<i>Item</i>	<i>Number of categories</i>	<i>Item difficulty (logits)</i>	<i>SE</i>	<i>Residual</i>	$\chi^2$	<i>Probability</i>
Comprehension	7	-0.48	0.07	0.31	2.88	0.09
Expression	7	-0.43	0.07	0.80	0.58	0.44
Social interaction	5	-0.20	0.09	0.89	0.56	0.45
Problem solving	4	0.94	0.08	-0.69	0.94	0.33
Memory	4	0.17	0.09	1.59	0.25	0.62

Bonferroni corrected  $P = 0.01$ 

## Discussion

In the present paper, the cross-cultural validity of motor items of FIM™ and the possibility to pool data from different countries in patients with spinal cord injury is evaluated using Rasch analysis. It is demonstrated that the number of categories presently used in FIM™ with data from routine clinical settings in the participating countries in Europe is not sustainable. Also after adjusting for such problems, items have different levels of difficulty across countries. Thus, the pooling of raw data is not recommended, casting doubts on the cross-

cultural validity of the instrument. However, it is demonstrated that by allowing some items to be unique for each country, this limitation can be overcome to a great extent. This allows generating estimates of ability for a patient, which can be confidently used to represent the same level of dependence regardless of the country of the patient. Unfortunately, the rescoring and splitting of items across countries is a very complex solution, and not ideal for clinical studies of activity limitation. A similar approach has been used for the motor items FIM™ in patients after stroke,<sup>28</sup> using eight split and five original common items as in the present study

of patients with SCI, although the common items were different. In SCI, four country specific items had to be deleted to get fit to the model, as well as two items common to all countries. In stroke, however, only three country-specific items had to be deleted.<sup>28</sup> To delete an item, as for example, the Eating and Grooming item in SCI, jeopardises the clinical use of the instrument as essential daily activity items are omitted. Thus, the relevance of the FIM™ for this patient group can be questioned, at least when comparing data from different countries. A comparison of data from different settings within the same country may show similar problems, although this is not addressed in the present report.

At least 30% of the patients were at the ceiling of the social cognitive scale in each country, with 61% at the extreme in Denmark. As a result, the data from approximately 60 patients were available for the Rasch analysis for each country, with consequent attendant problems of parameter estimation and precision of the item difficulty estimate. The large number of extreme patients raises questions about the validity of the social-cognitive items for this patient group. In addition, earlier studies have shown no relationships between results from a comprehensive, predominately motor-free, neuropsychological test battery and the results from the FIM™ social-cognitive items, also attributed to the ceiling effects in the FIM ratings.<sup>11</sup>

The approach used in this paper is one of identifying problems with cross-cultural validity through the analysis of DIF. DIF is a breach of the assumptions of unidimensionality,<sup>20</sup> but it has also been argued that DIF can be evaluated only if the conditions for fit to the Rasch model have been satisfied.<sup>29</sup> As it is possible to have data which fit the model, but which also display DIF, we have adopted a pragmatic approach whereby we have considered DIF as one possible contribution to misfit, and thus deliberately refitted data to the model after adjustment for DIF.

Many factors may have contributed to the cross-cultural variability observed in the data. There may be some difference in the reliability of the professionals who rate the FIM™, and in the translation of the FIM™ manual. Different manuals and training in different settings may have had an influence on the psychometric quality of the instrument, although all things being equal, those with formal FIM™ training has been shown to give more reliable ratings.<sup>30</sup> As data collection was restricted to scale items, age, and gender, we have not been able to analyse the impact of training or rater reliability. True cultural differences in the ways people dress or bathe may also contribute to the lack of invariance.<sup>31</sup> The influence of all these factors and the extent of their interaction are unknown. Therefore, the lack of information about differences between datasets and case mix (eg level of injury, time since injury, complete/incomplete injury, traumatic or nontraumatic cause) is a limitation of the study. However, it must be clearly stated that a requirement for pooling data from different countries is that the scale should be invariant across those countries. If the lack of invariance was

attributable to aspects such as case mix, level of training of raters, and other administrative factors, then this suggests that the scale would be unsuitable for such studies. The task of adjusting for so many effects would be onerous, given the example above where there was a loss of clinically relevant information in solving the problem at the crude country level.

In conclusion, cross-cultural data from the FIM™ motor scale for patients with spinal cord injury cannot be pooled in its raw form. In order to make comparison across countries, it is necessary to accommodate cultural differences in the measurement construct by making complex adjustments using the Rasch model. In measurement terms the scale still remains problematic. The calculation of change scores and use of parametric statistics on the raw score, between and within different countries, are inadvisable. This does not preclude, in any way, the use of the FIM instrument in clinical practice, though the analysis clearly highlights problems with the seven category response function.

## References

- 1 Meyers AR, Andresen EM, Hagglund KJ. A model of outcomes research: spinal cord injury. *Arch Phys Med Rehabil* 2000; **81**(Suppl 2): S81–S90.
- 2 Haigh R *et al*. The use of outcome measures in physical medicine and rehabilitation within Europe. *J Rehabil Med* 2001; **33**: 273–278.
- 3 Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; **25**: 3186–3191.
- 4 World Health Organisation. *The International Classification of Functioning, Disability and Health*. WHO: Geneva 2002.
- 5 Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil* 2000; **81**(Suppl 2): S21–S29.
- 6 Keith R, Granger C, Hamilton B, Sherwin F. The Functional Independence Measure: a new tool for rehabilitation. In: Eisenberg M, Grzesiak R (eds). *Advances in Clinical Rehabilitation*. New York: Springer Publishing Co 1987 pp 6–18.
- 7 Eastwood EA, Hagglund KJ, Ragnarsson KT, Gordon WA, Marino RJ. Medical rehabilitation length of stay and outcomes for persons with traumatic spinal cord injury – 1990–1997. *Arch Phys Med Rehabil* 1999; **80**: 1457–1463.
- 8 McKinley WO, Seel RT, Gadi RK, Tewksbury MA. Nontraumatic versus traumatic spinal cord injury: a rehabilitation outcome comparison. *Am J Phys Med Rehabil* 2001; **80**: 693–699; quiz 700, 716.
- 9 Putzke JD, Hicken BL, Richards JS. Race: predictor versus proxy variable. Outcomes after spinal cord injury. *Arch Phys Med Rehabil* 2002; **83**: 1603–1611.
- 10 Segal ME, Ditunno JF, Staas WE. Interinstitutional agreement of individual functional independence measure (FIM) items measured at two sites on one sample of SCI patients. *Paraplegia* 1993; **31**: 622–631.
- 11 Davidoff GN, Roth EJ, Haughton JS, Ardner MS. Cognitive dysfunction in spinal cord injury patients: sensitivity of the Functional Independence Measure subscales versus neuropsychologic assessment. *Arch Phys Med Rehabil* 1990; **71**: 326–329.

- 12 Grey N, Kennedy P. The Functional Independence Measure: a comparative study of clinician and self ratings. *Paraplegia* 1993; **31**: 457–461.
- 13 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.
- 14 Rasch G. *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press: Expanded ed. Chicago 1980.
- 15 Smith RM. Fit analysis in latent trait measurement models. *J Appl Meas* 2000; **1**: 199–218.
- 16 Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978; **43**: 561–573.
- 17 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.
- 18 Tennant A, Hillman M, Fear J, Pickering A, Chamberlain MA. Are we making the most of the stanford health assessment questionnaire? *Br J Rheumatol* 1996; **35**: 574–578.
- 19 Prieto L, Alonso J, Lamarca R, Wright BD. Rasch measurement for reducing the items of the Nottingham Health Profile. *J Outcome Meas* 1998; **2**: 285–301.
- 20 Angoff W. Perspectives on Differential Item Functioning Methodology. In: Holland P, Wainer H, (eds). *Differential Item Functioning*. Lawrence Erlbaum: Hillsdale, New Jersey 1993.
- 21 Tennant 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.
- 22 Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Measure* 2002; **3**: 205–231.
- 23 Linacre JM. Investigating rating scale category utility. *J Outcome Measurement* 1999; **3**: 103–122.
- 24 Orlando M, Marshall GN. Differential item functioning in a Spanish translation of the PTSD checklist: detection and evaluation of impact. *Psychol Assessment* 2002; **14**: 50–59.
- 25 Elashoff JD. *NQuery Advisor. Version 4.0. User's Guide*. Boston: Statistical Solutions, 2000.
- 26 Andrich D, Lyne A, Sheridan B, Luo G. *RUMM 2020*. RUMM Laboratory: Perth, 2003.
- 27 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; **310**: 170.
- 28 Lundgren-Nilsson 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.
- 29 Smith RM. A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modelling* 1996; **3**: 35–40.
- 30 Fricke J, Unsworth C, Worell D. Reliability of the Functional Independence Measure with occupational therapist. *Aus Occupational Ther J* 1993; **40**: 2–15.
- 31 Küçükdeveci AA, Yavuzer G, Ehan AH, Sonel B, Tennant A. Adaptation of the Functional Independence Measure for use in Turkey. *Clinical Rehabil* 2001; **15**: 311–319.