

LETTER TO THE EDITOR

COMMENTS TO DIFFERENTIAL ITEM FUNCTIONING OF THE FUNCTIONAL INDEPENDENCE MEASURE IN HIGH PERFORMING NEUROLOGICAL PATIENTS

Sir,

The article by AJ Dallmeijer et al. (1) is a carefully performed and well-written study, but its positives are vastly outweighed by its negatives. I offer the following rebuttal:

- In the study, the mean scores for the 3 groups of patients were more than 100 points out of the motor-plus-cognition maximum possible of 126 points. When the total score is between 100 and 110 points, patients no longer need assistance from another person. All but a few item scores were at the levels No. 6 or 7 ($6 \times 18 = 108$). Level 6 is “modified independence”, meaning that help from another person is not needed. Level 7 is “complete independence” of help from another person or device. The FIM™ instrument was designed for use with patients who need assistance from another person in order to measure “burden of care”. Outpatients and those who have completed a course of inpatient rehabilitation usually do not require assistance or require only minimal assistance from another person for personal care activities on a daily basis.
- Published studies have shown the relationships between total FIM™ instrument scores and burden of care as measured by the amount of time in minutes that a helper is needed on a daily basis. Thus, the score in terms of its conversion into “burden of care” time needed for assistance from another person, and possibly related costs, accounts for the underlying validity of the FIM™ instrument. The conversion in the *mid-range* of the FIM™ instrument is equivalent to saving 3–5 minutes of help per day with each gain of one FIM™ instrument point. Thus, gaining 10 points amounts to saving the helper 30–50 minutes per day. With awareness that the FIM™ instrument is designed to measure the burden of care in terms of time required of the helping person, then the notion of a “ceiling effect” is not appropriate. Hence, when performing studies to examine the foundations of the FIM™ instrument, it is imperative that all 7 scoring levels will be sufficiently represented, especially levels 1–5. The most reliable method for doing this is to examine both admission (lower) scores and discharge (higher) scores together. To do otherwise skews the data in one direction or the other. Taking stroke as an example, a motor score of 76 probably would yield 10 item scores of 6 or 7 and 3 item scores of 4 or 5, while a cognition score of 31 probably would not yield any item scores below 6. Further, the authors do not mention removal of individual patients who misfit in order to reduce another source of distortion of the analysis before proceeding to make comparisons.

- References regarding the correlation of FIM™ instrument scoring to minutes of assistance are given below (2–7).
- While it is stated that “The investigation was conducted by having trained clinical researchers (physiatrists) collect FIM™ item scores by direct observation of and interviews with the patients or by interviewing proxies or caregivers,” there is no indication that the physiatrists had passed a FIM™ instrument mastery test administered by Uniform Data System for Medical Rehabilitation (UDSMR) in order to demonstrate that the clinicians can rate patients accurately. Further, determining ratings by interviewing proxies or caregivers rather than by direct observation requires skills that may not have been present.
- Finding that bladder and bowel items misfit is common. The most likely explanation is that bladder and bowel management are partly voluntary and partly involuntary (as autonomic functions). Therefore, these items often do not have fixed locations in the rating hierarchy among the voluntary items, particularly for subjects who have neurological conditions or are older. From a functional point of view, one must consider that a patient with a “mild” bladder impairment may have that difficulty rated as moderate or marked if a concurrent mobility limitation or environmental barrier causes the patient to require assistance in toileting or in bladder management.
- I disagree that Rasch estimates are more accurate indicators of person ability (or change in ability) at either the lower or higher end of the scoring range than the raw FIM™ instrument scores because the Rasch model is designed to assume infinity at either end of a measure. It is the investigator who imposes an arbitrary limit at either end of a measure for the sake of clinical utility.
- Differential item functioning (DIF) analysis is a test for whether an item or items occupy different or unexpected positions along the hierarchy that makes up a measure. Yi Du (8) identifies 3 criteria for judging DIF: (i) Is there statistical significance? (ii) Are there substantive implications of a practical consequence that lead to different interpretations of the results? (iii) Is the DIF associated with real differences in subjects tested or is it due to an accident of sampling? If all 3 criteria are met in a test situation, then Du (8) suggests that the remedy for having one item operating like 2 different items is to adjust the item in reference and in focal groups, accordingly. On the other hand, there are myriad reasons for DIF not systematically accounted for in this article, including: age grouping; timing of measurement; careless, uninformed or incompletely trained raters; altered environment; disease conditions with

varying underlying pathophysiologies (such as inclusion of stroke and traumatic brain injury (TBI) with multiple sclerosis (MS) in the sample without differentiation); imposed restraints or else enhanced performance induced by functional aids such as using a wheelchair; or other reasons.

- Several studies have been published illustrating that a certain degree of DIF is characteristic of the types of patients commonly treated as rehabilitation inpatients. Granger & Lin (9) carried out a study in which approximately 1000 patients from each of 19 impairment groups were studied from 1994 UDSMR data. To assure distribution of low, middle, and high raw FIM™ instrument scores, admission, discharge, and follow-up records were included. Results from each analysis were plotted and visually inspected. When the hierarchical pattern of functional activities was substantially similar to that from another impairment group, the data from the 2 groups were combined. This process was continued until 5 distinct hierarchical patterns were identified for the FIM™ instrument motor items: (i) brain dysfunction (including TBI, non-traumatic brain injury, and stroke); (ii) orthopedic and related conditions; (iii) pain conditions; (iv) spinal cord dysfunction: walkers (for patients with traumatic or non-traumatic spinal cord conditions who were walking at discharge); and (v) spinal cord dysfunction: wheelchair users (for patients with traumatic or non-traumatic spinal cord conditions who used a wheelchair at discharge. Individual patients with MS might fit into group 1, 4 or 5. For the FIM™ instrument cognition items, there were 2 patterns: (i) stroke with right body hemiparesis (left cerebral hemisphere involvement) or (ii) all others. (Important DIF of cognition items is noted when right- and left-sided strokes are analyzed separately.) Once these primary and cognitive patterns were identified, item logit values were anchored using the values calculated from the relevant combined group. Using anchored values, admission, discharge, and follow-up item scores for each impairment group were subjected to separate Rasch analyses.

My impression is that this study is an example of test bias (in contrast to item bias) because high functioning patients are inappropriate for this type of study of the FIM™ instrument. The instrument is designed to measure burden of care. Penfield & Lam (10) state that test bias occurs when performance on the test requires sources of knowledge different from those intended to be measured, causing the test scores to be less valid for a particular group. The goal of DIF detection is to determine whether the performance on an item differs between groups of subjects having the same estimated ability. Also, DIF detection attempts to disentangle the effects of fairness (from a test point of view) from ability level on the between-group differences in item performance. In test bias, not only must it be shown that subjects of equivalent ability perform differently on an item, but that the observed difference in item performance can be attributed to some property of the item that is unrelated to the construct intended to be measured by the test. Thus, only members of 2 groups that are at the same level of ability are to be compared. In

typical investigations into item bias, *statistical* DIF (which is the common use) is first assessed, then, for those items displaying statistical DIF, *substantive* DIF is assessed through consideration of item content. Only when both statistical and substantive DIF exist is the item considered biased. In addition, statistical DIF methodology must be used in combination with construct validity investigations in order to ensure the proper assessment of item bias. Several real reasons for DIF in the sample were overlooked, such as lack of evidence that the raters were trained and attained accuracy in using the FIM™ instrument, which could have resulted in reduced reliability. Criteria proposed by Du (8) for when to adjust for DIF were not addressed. Further, all levels of answer categories were insufficiently represented in analysis, and collapsing of categories 1–5 has combined with this to produce flawed inferences.

REFERENCES

1. Dallmeijer AJ, Dekker J, Roorda LD, Knol DL, Balen B van, Groot V de, et al. Differential item functioning of the Functional Independence Measure in higher performing neurological patients. *J Rehabil Med* 2005; 37: 346–352.
2. Granger CV, Cotter AC, Hamilton BB, Fiedler RC, Hens MM. Functional assessment scales: a study of persons with multiple sclerosis. *Arch Phys Med Rehabil* 1990; 71: 870–875.
3. Granger CV, Cotter AC, Hamilton BB, Fiedler RC. Functional assessment scales: a study of persons after stroke. *Arch Phys Med Rehabil* 1993; 74: 133–138.
4. Granger CV, Divan N, Fiedler RC. Functional assessment scales: a study of persons after traumatic brain injury. *Am J Phys Med Rehabil* 1995; 74: 107–113.
5. Disler PB, Roy CW, Smith BP. Predicting hours of care needed. *Arch Phys Med Rehabil* 1993; 74: 139–143.
6. Corrigan JD, Smith-Knapp K, Granger CV. Validity of the Functional Independence Measure for persons with traumatic brain injury. *Arch Phys Med Rehabil* 1997; 78: 828–834.
7. Hamilton BB, Deutsch A, Russell C, Fiedler RC, Granger CV. Relation of disability costs to function: spinal cord injury. *Arch Phys Med Rehabil* 1999; 80: 385–391.
8. Du Y. When to adjust for Differential Item Functioning. *Rasch Measurement Transactions* 1995; 9: 414.
9. Granger CV, Linn RT. Biologic patterns of disability. *Journal of Outcome Measurement* 2000; 4: 595–614.
10. Penfield RD, Lam TCM. Assessing differential item functioning in performance assessment: review and recommendations. *Educational Measurement: Issues and Practice*. Fall 2000; 19: 5–15.

Submitted March 16, 2006; accepted April 29, 2006

Carl V. Granger, MD

Professor and Chairman Emeritus,
Department of Rehabilitation Medicine,
University at Buffalo,
The State University of New York,
Buffalo, New York USA, and
Executive Director,
Uniform Data System for Medical Rehabilitation
270 Northpointe Parkway, Suite 300,
Amherst, New York 14228 USA.
E-mail: cgranger@udsmr.org

RESPONSE TO LETTER TO THE EDITOR BY CARL V. GRANGER

In his *Letter to the Editor* Dr Granger raises some interesting points of discussion about our paper on differential item functioning of the FIM™ in higher functioning patients with neurological disorders. We appreciate his comments on our paper and like to make some remarks and explain our point of view.

The purpose of the study was to examine differential item functioning in higher performing neurological patients. The observations from this study cannot be generalized as such towards patients with lower levels of functioning; the study was aimed at higher performing patients, and the results should be interpreted within this context. We have elaborated on this in the discussion.

Our study demonstrates some limitations of the FIM™ (misfit of some FIM™ items, disordered thresholds, and occurrence of DIF) that should be acknowledged when using the instrument in these specific patient groups, and when different groups are pooled. Re-scoring of answering categories and adjusting items for DIF may be used to improve the applicability of the instrument in these groups. Further, we like to point out that a recent study in stroke patients at admission of rehabilitation (1) reported similar limitations. The distinct hierarchical patterns for the FIM™ items that are identified for different impairment groups, as illustrated by Dr Granger in his letter, also show that scores of different diagnostic groups are not directly comparable.

Since the FIM™ is widely used in a large number of diagnostic groups, in different phases of the rehabilitation process, and in different countries or cultures, we think that it is of paramount

importance to explore the applicability of the FIM™ in well-defined groups and acknowledge possible limitations. Further research into different diagnostic groups may lead to recommendations to improve the applicability and validity of the instrument in specific groups.

REFERENCE

1. Lundgren-Nilsson A, 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.

Annet J. Dallmeijer,¹ Joost Dekker,¹

Leo D. Roorda,² Dirk L. Knol,³

Bianca van Baalen,⁴ Vincent de Groot,¹

Vera P.M. Schepers⁵ and Gustaaf J. Lankhorst¹

¹Department of Rehabilitation Medicine and

²Jan van Bremen Institute, Center for Rheumatology and Rehabilitation Amsterdam,

³Department of Clinical Epidemiology and Biostatistics, VU University Medical Center,

⁴Departments of Rehabilitation Medicine, Erasmus Medical Center Rotterdam,

⁵Department of Rehabilitation Medicine, University Medical Center Utrecht,

The Netherlands.

E-mail: a.dallmeijer@vumc.nl