DISABILITY AND REHABILITATION, 1995; VOL. 17, NO. 1, 10–14

The Functional Independence Measure: a comparative validity and reliability study

D. KIDD, G. STEWART, J. BALDRY, J. JOHNSON, D. ROSSITER, A. PETRUCKEVITCH and A. J. THOMPSON

Accepted for publication: October 1993

Keywords Functional Independence Measure, Barthel Index, neurological disability

Summary

The majority of measurement scales used to evaluate outcome in rehabilitation are ordinal in nature and consequently statistically valid assessments of change are difficult to make. The Functional Independence Measure (FIM) can be weighted to possess interval properties, potentially allowing more accurate analysis of change. In this study the FIM was compared to the Barthel Index (BI) to determine its validity, reliability and ease of use in two groups of 25 patients undergoing neurorehabilitation. The FIM was considered to be more valid than the BI, and equally reliable in the assessment of disability. When the two disability scores were compared using subjective and objective assessment the agreement between them was comparable, although neither was high.

Introduction

In order to evaluate accurately the efficacy of any therapeutic practice it is necessary to use appropriate measuring tools which are both valid and reliable.[1,2] This is of particular importance in rehabilitation, which is labour-intensive and is often carried out over a prolonged period of time. The difficulties inherent in clinical evaluation, particularly in relation to the assessment of change, have resulted in the development of a wide variety of measurement scales which attempt to give a numerical notation to the clinical situation. Such measurement of clinical change is crucial to the assessment of outcome, benefit and cost-effectiveness.

Most scales used are ordinal in nature; that is they show a rank order, but it cannot be assumed that the steps on the scale are equal. It follows that it cannot be concluded that a score of 10 is twice that of 5, and this inequality considerably limits statistical analysis and interpretation. Hence it is not correct to use the change in an ordinal scale to produce a numerical notation of corresponding clinical change unless all patients start at the same point. Intervals scales, on the other hand, are ranked in equal proportions, allowing comparison of change, even when starting at different points in the scale, and therefore providing a more objective way of quantifying functional change in groups of disabled patients.

The Barthel Index[3] (BI) (Appendix 1) is an ordinal scale with a range of 0 to 20 which has been used as a measure of disability in clinical research for many years. Previous studies have stated that it is valid, relatively reliable but not very sensitive to change.[4–6] The Functional Independence Measure[7–9] (FIM) (Appendix 2) was devised in 1984 by a workshop developed by the American Congress for Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation to be used as a universal assessment tool in the Uniform Data System for Medical Rehabilitation (UDS[10]). It was developed in order to provide a more comprehensive assessment of disability not only of self-care activities and mobility but also communication and cognitive function. It has a range of 13–126. Initially it, too, was ordinal but has since undergone several refinements — the most notable of which is the weighting of individual scores using Rasch analysis,[11] which utilizes log odds ratios of probability to make the intervals between individual steps the same, and thus allows the FIM to be used as an interval scale.[12] It comprises 18 items; 13 motor items (self-care and locomotion) and five cognitive items (communication and cognition), each of which is scored on a seven-point scale. The motor and cognitive scores have been separated, and each score now has a range of 0–100. The interrater reliability is said to be high.[13]

Giving the FIM interval properties allows it to be used

Correspondence to: Dr A. J. Thompson, Neurorehabilitation Unit, Institute of Neurology, National Hospital, Queen Square, London WC1N 3BG, UK.

0963-8288/95 $10·00

statistically to measure change in disability, and so it can more accurately be used as an objective assessment of change both in rehabilitation research and in clinical trials.

The aims of this study were: (a) to compare the validity and ease of use of the FIM with the BI as an index of disability at a single point in time, (b) to compare the rate of change of both scales over a period of inpatient rehabilitation, and (c) to assess the reliability of both scales when assessed using a single scorer and a multidisciplinary team scoring the scale with subjective and objective information respectively.

Methods

VALIDITY STUDY

Twenty-five consecutive patients admitted to the neurorehabilitation unit were studied. FIM(motor) and BI were each scored within 3 days of admission by a multidisciplinary team comprising a doctor, an occupational therapist, a physiotherapist, a speech therapist and the primary nurse for the patient concerned. Measurement was based on the information obtained by individual members of the team having assessed the patient. Within 3 days of planned discharge the two scores were repeated in the same manner without access to the admission scores. The FIM(motor) scores were then adjusted to the Rasch analysed weighting table.[12]

RELIABILITY STUDY

The following 25 patients admitted were also scored on the FIM(motor) and BI. In this study the medical member of the team (D.K.) interviewed each patient directly without prior knowledge of their true ability from therapists; thus the score was based solely on patient report. The same patients were scored independently by the remainder of the team using the same methods as described previously. The scores were repeated within 3 days of discharge.

STATISTICAL ANALYSIS

For the validity study the kappa statistic κ was used to measure the degree of agreement between the two scores, and includes a correction for any agreement which may occur by chance. In the reliability study the method proposed by Bland and Altman[14] was employed to estimate the precision of, and agreement between, the measurements made by the different observers.

Results

VALIDITY STUDY

Twelve patients had multiple sclerosis (MS), six had had strokes, one a central nervous system vasculitis and two each had brain and spinal cord tumours and head injury. The median length of stay was 23 days. The median (range) BI was 12 (2–20) on admission, 19 (4–20) on discharge with a median change of 3 (0–13). The median FIM(motor) was 54 (30–100) on admission and 71 (33–100) on discharge, with a median change of 7 (−4–34). Twenty (80%) patients improved on the BI, 21 (84%) improved on the FIM. Two patients were unchanged on either score.

When the individual components of each score were examined (the 13 components of the FIM were necessarily reduced to the same 10 components of the BI) 14% of the scored components changed on the FIM but not on the BI, whereas only 2% showed change on the BI but not the FIM. Thirty per cent changed on both scores. The disparity between FIM and BI was seen in all 10 components; the most frequently occurring were dressing, grooming, bathing, transfers and ability to climb stairs.

Table 1 shows the degree of agreement between the two scores as indicated by the kappa statistic κ. This shows that, while the agreement between the two scores was reasonable on admission and discharge it was less so for change in score.

RELIABILITY STUDY

Twelve patients had MS, three had had strokes, two each had head injury, brain tumour, cord tumour and movement disorder, one had Guillain-Barré syndrome and one had had encephalitis. The medium length of stay was 22 days.

The median (range) BI scored by the team and by the single observer was 15 (3–20) and 16 (3–20) respectively on admission, and on discharge 18 (4–20) and 19 (6–20) respectively, with a median change of 2 (0–9) and 1 (0–9). The median FIM(motor) scored by the team and by the single observer was 63 (24–83) and 66 (31–83) respectively on admission and 73 (31–90) and 68 (35–100) on discharge, with a median change of 8 (−1–37) and 4 (−3–35).

The objective measurement showed improvement in 76% with the BI and 92% with the FIM, whereas the subjective assessment showed improvement in 52% and 72% respectively. Table 2 shows the mean difference $\bar{d}$ between each set of scores and the 95% limits of agreement of scores of each scale. This shows quite wide

Table 1. Tables showing incidence of agreement between the two scores using a cut-off of 10 and 50 for BI and FIM respectively, with the kappa statistic of agreement and 95% confidence limits for BI and FIM(motor) scores of 25 patients in study[1]

	Admission				*Discharge*				*Change*			
		BI 0–10	BI 11–20			BI 0–10	BI 11–20			BI 0–10	BI 11–20	
FIM	0–50	11	1	12	0–50	5	0	5	0–50	18	0	18
	51–100	0	13	13	51–100	1	19	20	51–100	2	5	7
		11	14	25		6	19	25		20	5	25
Kappa statistic		0·92				0·88				0·78		
95% confidence limits		0·77–1·0				0·66–1·0				0·49–1·0		

Table 2. Mean difference and 95% limits of agreement between the two methods of scoring for BI and FIM(motor) scores of 25 patients in study[2]

	BI (admission)	*BI (discharge)*	*BI (change)*	*FIM (admission)*	*FIM (discharge)*	*FIM (change)*
Mean difference ($\bar{d}$)	0·8	0·44	1·24	2·56	0·64	3·20
95% limits of agreement	−4·72–3·12	−2·02–2·9	−2·19–4·67	−15·3–10·18	−16·8–18·08	−6·67–13·07

variations between the two methods of scoring which are proportionately comparable in both the FIM and the BI.

Discussion

In order to be an effective measurement tool which is useful in clinical practice a scale should possess the following fundamental properties: (a) validity: that is, it actually measures what it purports to measure; (b) reliability: that the measurement is repeatable and reproducible when measured by single and different observers; (c) sensitivity; (d) specificity; (e) ease of use and (f) non-specialist dependence. It is impossible to assess accurately sensitivity and specificity using these scales, since there exists no gold standard with which to compare them. Ease of use is of particular importance in clinical practice in order not to be time- and personnel-consuming, although in research greater accuracy is often achieved at the expense of time. In this study a comparison of two disability scales has been made addressing the validity, ease of use and reliability using different scoring methods.

VALIDITY

The Barthel Index is said to be valid[4,5] but is flawed in that 'ceiling' and 'floor' effects exist,[15] meaning that a spectrum of disability exists at the minimum and maximum scores which may not be identified if it changes. However, its 10 component parts provide an account of the functional status of the patient with respect to personal care and mobility. The Functional Independence Measure was developed in order to provide a measure of disability not only in terms of activities of daily living (ADL) but also communication and cognitive function. In expanding the motor components to 13 parts it provides a more comprehensive assessment of function than the BI. Whilst feeding, grooming, bathing and toileting are included in both, the FIM expands dressing and transfer activities to give a more detailed appraisal of the patient's function; most patients have greater difficulty dressing their bottom half, and the degree of help required will dictate how much time is spent dressing and undressing the patient by a carer. Similarly if the patient can transfer independently from wheelchair to toilet, but not into the bath or shower, then the degree of independence is much greater than for someone who requires help with all transfers. The locomotion component is also more valid in that it scores ability to move no matter by what means, while the BI penalizes patients who are wheelchair-bound but entirely independent. Furthermore the FIM ranks seven levels of severity, providing a more accurate assessment of ability in each component and increasing the potential to detect change should change occur. Finally the ability to identify and score problems with communication and cognition is particularly important, since these may be relevant to other functional limitations, and must be considered when planning rehabilitation programmes. Neither scale, however, scores instrumental ('extended') ADL tasks such as kitchen activities.

RELIABILITY

In the first study the κ statistic for both scores on admission and on discharge was similar, although lower for the assessment of change. It may be that this is due to the fact that the FIM is more sensitive in identifying change; when all the scores are analysed 84% of patients were noted to improve on the FIM against 67% on the BI, and when individual component parts are examined 14% showed change on the FIM alone, whereas only 2% showed change on the BI without a corresponding change on the FIM.

In the second study the reliability of the BI was comparable to the FIM (Table 2). The subjective assessments tended to be scored at higher levels than the objective ones, perhaps reflecting the fact that patients tend to underestimate their disability.[16] Furthermore the charge in scores was smaller on subjective assessment than when objectively assessed by the multidisciplinary team. This has important implications for the use of such scales in research and clinical practice. It has been stated that the BI can be scored by patient report even on the telephone, but our data do not hold this to be true. The variation in BI ranged from -3 to $+6$, and was proportionately similar to that of the FIM. The degree of agreement between BI scores was similar to that noted previously.[6,17] The results do, however, suggest that repeated measurements should be carried out by the same observers each time.

EASE OF USE

The BI is easy and quick to perform and requires little expertise in its use.[4] On the other hand the FIM requires some training and experience in the scoring system, takes longer to complete and frequent reference to the scoring booklet is needed, although with practice it is possible to score the FIM quite quickly.

In summary our preliminary experience of the FIM suggests that it is a valid and useful scale with reliability similar to that of the BI. Furthermore the use of Rasch analysis transforms the FIM into an interval scale (although the individual components remain ordinal), permitting more meaningful statistical analysis to be made of patient groups. The greater choice of scores allows more accurate representation of true abilities. It is, however, more time-consuming to score and requires experience in its use. The results show that it is important to score patients using objective measurements rather than by patient report, since the latter underestimates disability and identifies less change in function subsequently. The studies suggest that the FIM may be considered superior to the existing ADL scales and has a role in both research and in clinical practice to assess patients' functional status and to monitor results from therapeutic intervention or disease progression.

Acknowledgements

Dr Kidd is supported by a grant from the Brain Research Trust. The authors thank all members of staff at the neurorehabilitation unit for their involvement, and Dr R. Howard for useful comments.

References

1 LaRocca NG. Statistical and methodologic considerations in scale construction. In: Munsat TL (ed.) *Quantification of Neurologic Deficit*. Boston: Butterworth, 1989; 49–66.
2 McKenzie CR, Charleson ME. Standards for the use of ordinal scales in clinical trials. *British Medical Journal* 1986; **292**: 40–43.
3 Mahoney FI, Barthel DW. Functional evaluation — the Barthel Index. *Maryland State Medical Journal* 1965; **14**: 61–65.
4 Wade DT, Collin C. The Barthel Index: a standard measure of physical disability? *International Disability Studies* 1988; **10**: 64–67.
5 Gresham GE, Phillips TF, Labi MLC. ADL status in stroke: relative merits of three standard indexes. *Archives of Physical Medicine and Rehabilitation* 1980; **61**: 355–358.
6 Collin C, Wade DT, Davis S, Horne V. The Barthel Index: a reliability study. *International Disability Studies* 1988; **10**: 61–63.
7 Granger CV, Hamilton BB, Keith RA, Zielezny M, Sherwin FS. Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation* 1986; **1**: 59–74.
8 Keith RA, Granger CV, Hamilton BB, Sherwin FS. The Functional Independence Measure: a new tool for rehabilitation. In Eisenberg MG, Grzesiak RC (eds) *Advances in Clinical Rehabilitation*, vol. 1. New York: Springer-Verlag, 1987; 6–18.
9 Minaire P. La Mesure d'Indépendance Fonctionelle (MIF): historique, présentation, perspectives. *Journal Réadaptation Médicine* 1991; **11**: 168–174.
10 Hamilton BB, Granger CV, Sherwin FS, Zielezny M, Tasman JS. A uniform national data system for medical rehabilitation. In Fuhrer MJ, (ed.) *Rehabilitation Outcomes: Analysis and Measurement*. Baltimore: Brookes, 1987; 137–147.
11 Andrich D. *Rasch Models of Measurement*. Beverly Hills: Sage Publications, 1988.
12 Fisher AG, Bryze KA, Granger CV, Haley SM, Hamilton BB, Heinemann AW, Jalayerian JK, Linacre JM, Ludlow LH, McCabe MA, Wright BD. Applications of Rasch analysis to the development of functional assessments. *International Journal of Education Research* 1993 (In press).
13 Hamilton BB, Laughlin JA, Granger CV, Kayton RM. Interrater agreement of the seven level Functional Independence Measure (FIM). *Archives of Physical Medicine and Rehabilitation* 1991; **72**: 790 (abstract).
14 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–310.
15 Applegate WB, Blass JP, Williams TF. Instruments for the functional assessment of older patients. *New England Journal of Medicine* 1990; **322**: 1207–1214.
16 Rubenstein LZ, Schairer C, Wieland GD, Kane R. Systematic biases in functional status assessment of elderly adults: effects of different data sources. *Journal of Gerontology* 1984; **39**: 386–391.
17 Loewen SC, Anderson BA. Reliability of the Modified Motor Assessment scale and the Barthel Index. *Physical Therapy* 1988; **68**: 1077–1081.

Appendix 1: The Barthel Index

Bowels	0 — incontinent 1 — occasional accident 2 — continent
Bladder	0 — incontinent or catheterized and unable to manage 1 — occasional accident 2 — continent
Grooming	0 — needs help 1 — independent
Toilet use	0 — dependent 1 — needs some help but can do something 2 — independent
Feeding	0 — unable 1 — needs help cutting, spreading butter, etc. 2 — independent
Transfer	0 — unable 1 — major help (one or two people) 2 — minor help (verbal or physical) 3 — independent
Mobility	0 — immobile 1 — wheelchair-independent including corners, etc. 2 — walks with the help of one person (verbal or physical) 3 — independent (but may use an aid)
Dressing	0 — dependent 1 — needs help, but can do about half unaided 2 — independent
Stairs	0 — unable 1 — needs help (verbal, physical, carrying aid) 2 — independent up and down
Bathing	0 — dependent 1 — independent

2: The Functional Independence Measure

FIM(motor)	*FIM(cognitive)*
Self-care	**Communication**
A. Feeding	N. Comprehension
B. Grooming	O. Expression
C. Bathing	
D. Dressing upper body	**Social cognition**
E. Dressing lower body	P. Social interaction
F. Toileting	Q. Problem solving
	R. Memory
Sphincter control	
G. Bladder management	
H. Bowel management	
Mobility	
Transfer:	
I. Bed, chair, wheelchair	
J. Toilet	
K. Tub, shower	
Locomotion	
L. Walk/wheelchair	
M. Stairs	
Independence	7 — complete independence (timely, safely) 6 — modified independence (device)
Modified dependence:	5 — supervision 4 — minimal assistance (subject 75% +) 3 — moderate assistance (subject 50% +)
Complete dependence:	2 — maximal assistance (subject 25% +) 1 — total assistance (subject 0% +)