

Measuring change in disability after inpatient rehabilitation: comparison of the responsiveness of the Barthel Index and the Functional Independence Measure

J J M F van der Putten, J C Hobart, J A Freeman, A J Thompson

Abstract

Background—The importance of evaluating disability outcome measures is well recognised. The Functional Independence Measure (FIM) was developed to be a more comprehensive and “sensitive” measure of disability than the Barthel Index (BI). Although the FIM is widely used and has been shown to be reliable and valid, there is limited information about its responsiveness, particularly in comparison with the BI. This study compares the appropriateness and responsiveness of these two disability measures in patients with multiple sclerosis and stroke.

Methods—Patients with multiple sclerosis (n=201) and poststroke (n=82) patients undergoing inpatient neurorehabilitation were studied. Admission and discharge scores were generated for the BI and the three scales of the FIM (total, motor, and cognitive). Appropriateness of the measures to the study samples was determined by examining score distributions, floor and ceiling effects. Responsiveness was determined using an effect size calculation.

Results—The BI, FIM total, and FIM motor scales show good variability and have small floor and ceiling effects in the study samples. The FIM cognitive scale showed a notable ceiling effect in patients with multiple sclerosis. Comparable effect sizes were found for the BI, and two FIM scales (total and motor) in both patients with multiple sclerosis and stroke patients.

Conclusion—All measures were appropriate to the study sample. The FIM cognitive scale, however, has limited usefulness as an outcome measure in progressive multiple sclerosis. The BI, FIM total, and FIM motor scales show similar responsiveness, suggesting that both the FIM total and FIM motor scales have no advantage over the BI in evaluating change.

(*J Neurol Neurosurg Psychiatry* 1999;66:480-484)

Keywords: Barthel Index; Functional Independence Measure; responsiveness; rehabilitation

Measuring the effectiveness of clinical interventions by using standardised measurement instruments is now widely accepted as being

central to good clinical practice.¹ As the number of potential healthcare interventions has increased disproportionately to the health-care budget, pressure has been put on services to show that they provide high quality care that is cost effective.

Measuring the outcome of healthcare interventions is a central component of determining therapeutic effectiveness and, therefore, the provision of evidence-based health care. However, information generated by outcome studies is only meaningful if the measures used are clinically useful and scientifically sound.² Consequently, it must be shown that instruments measure the outcome under study in a way that is reliable and valid. In addition, instruments used for evaluative studies must also be shown to be able to detect clinically significant change in the outcome measured. This property is known as responsiveness.³

As rehabilitation is a labour intensive and costly intervention, evaluating its therapeutic effectiveness is particularly important. Although some studies have shown that rehabilitation is beneficial,⁴⁻⁹ there is no consensus as to which outcomes should be measured. Rehabilitation aims to improve various aspects of a patient's life—for example, disability, handicap, and quality of life—and ideally these should all be included in the outcome assessment.¹⁰ Despite a move towards quality of life and patient based outcome measures, observer rated generic measures of disability are still widely used.¹¹ The skills involved in self care and mobility are assumed to be basic to higher levels of functioning,¹² thus improvements in disability are likely to have considerable impact on a person's level of handicap and health related quality of life.

The Barthel Index (BI) and the Functional Independence Measure (FIM) are probably the most widely used generic disability measures. The BI was developed in 1955 as a simple index of independence useful in scoring disability.¹³ However, it was regarded as being too crude, too simple, and not responsive enough to evaluate disability outcomes in rehabilitation. Consequently, the FIM was developed between 1984 and 1987. The specific aims of the developers of the FIM were to produce an instrument that provided comprehensive and “sensitive” disability measurement.¹⁴ The FIM contains more items than the BI, includes cognitive items, and has more response categories.

Institute of Neurology,
Queen Square, London
WC1N 3BG, UK

J J M F van der Putten
J C Hobart
J A Freeman
A J Thompson

Correspondence to:
Professor A J Thompson,
Institute of Neurology,
Queen Square, London
WC1N 3BG, UK. Telephone
0044 171 837 3611 ext
4152; fax 0044 171 813
6505; email
athompson@ion.ucl.ac.uk

Received 3 March 1998 and
in revised form
5 November 1998
Accepted 11 November 1998

Although both instruments have evidence of reliability and validity,¹⁵⁻²⁰ there is only limited information about their responsiveness. Wade *et al* have suggested that the responsiveness of the BI is adequate for clinical purposes but may be limited in the context of research.¹⁵ Conceptually, it is argued that the FIM is more responsive than the BI.^{12 17 21} However, empirical data supporting this claim are limited, and those which exist use suboptimal methodology.¹⁸ Standard techniques for the assessment of responsiveness, such as the application of effect sizes, have not been used previously.

The aim of this study is to compare the appropriateness and responsiveness of the BI and the FIM in patients with multiple sclerosis and stroke patients receiving inpatient rehabilitation.

Methods

STUDY SAMPLE

Patients with multiple sclerosis and stroke patients who were admitted to the neurorehabilitation unit (NRU) of the National Hospital for Neurology and Neurosurgery between 1994 and 1997 were studied. These diagnostic groups were studied because they have different clinical disease courses (acute *v* chronic). The NRU is an 18 bed unit specialising in intensive, individually tailored, goal oriented rehabilitation of patients with neurological disorders.²² Patients are selected for admission to the NRU if they have the physical potential to actively participate in an intensive rehabilitation programme; the cognitive ability to carry over learned skills into functional tasks; and require input from at least two disciplines other than medical and nursing staff. Patients were excluded from this study if their duration of stay was less than 7 days.

MEASURES

The BI is a 10 item instrument measuring disability in terms of a person's level of functional independence in personal activities of daily living.¹⁶ It is rated from observation and has two items on a two point scale, six items on a three point scale, and two items on a four point scale. Item scores are summed to generate a total score (0=minimum independence; 20=maximum independence). The BI is user friendly and multiple studies support its reliability and validity.^{15 16 23}

The FIM is an 18 item instrument measuring a person's level of disability in terms of burden of care.¹⁴ It was developed specifically to measure functional outcomes of rehabilitation.²⁰ The developers recommend that the FIM is rated from patient observation by the consensus opinion of a multidisciplinary team. Each item is rated from 1 (requiring total assistance) to 7 (completely independent). Three independent FIM scores can be generated by summing item scores: a total score (FIM total: 18 items), a motor score (FIM motor: 13 items), and a cognitive score (FIM cognitive: 5 items). Multiple studies support the reliability and validity of FIM scales.^{18-21 24 25}

The expanded disability status scale (EDSS) is a multiple sclerosis specific, neurologist rated, index grading disease severity from 0 (normal neurological examination) to 10 (death due to multiple sclerosis) in 20 steps.²⁶ Rating is based on the medical history and the neurological examination. Although the EDSS has been heavily criticised,²⁷⁻²⁹ it remains the most widely used measure for multiple sclerosis due to the absence of well evaluated superior alternatives. Evidence supports the reliability and validity of the EDSS.³⁰

PROCEDURE

Patients referred for neurorehabilitation were assessed by a senior multidisciplinary team consisting of a neurologist, clinical nurse specialist, occupational therapist, and physiotherapist. Patients whom it was considered would benefit from inpatient neurorehabilitation had an admission date booked. On admission to the NRU patient characteristics were recorded along with disease severity (EDSS) in the multiple sclerosis group. For all patients, disability measures (BI and FIM) were rated within 96 hours of admission to, and within 48 hours of discharge from the NRU by consensus opinion of a treating multidisciplinary team.

STATISTICAL ANALYSIS

Appropriateness

Appropriateness attempts to define whether the range of disabilities in a study sample is similar to the range of disabilities covered by an instrument. In this study appropriateness was assessed by examining score ranges, means, SDs, and floor and ceiling effects for the BI and three FIM scales. Mean scores indicate the central tendency of the group, ideally these should lie near the midpoint of the scale range. Sample range and SD indicate the extent to which an instrument demonstrates variability in the study sample. The greater the variability detected the better an instrument discriminates between subjects. Floor and ceiling effects, calculated as the percentage of the sample scoring the minimum and maximum possible scores respectively, indicate the extent that scores cluster at the bottom and top of the scale range. Floor and ceiling effects represent a limited ability to discriminate between subjects. When an instrument measures a restricted range of health status floor and ceiling effects indicate that the range of disability measured by the scale is less than the range of disability occurring in the study sample. Floor and ceiling effects exceeding 20% are considered to be significant.³¹

Responsiveness

Responsiveness is defined as the ability of a measure to detect clinically important change in the outcome of interest.³² In this study responsiveness was determined using an effect size calculation, defined as mean change score (discharge minus admission) divided by the SD of admission (pretreatment) scores.³³ Effect sizes indicate, in SD units, the magnitude of change undergone by an instrument between two points in time. Therefore, the greater the

Table 1 Characteristics of patients with multiple sclerosis and stroke patients

Disease group	n	Male sex (%)	Age (y)		DOS (days)			Admission EDSS score range
			Mean (SD)	Range	Mean (SD)	Range	Median (SD)	
Multiple sclerosis	201	30.8	45 (11.2)	22-73	21.4 (11.0)	10-100	7.0 (1.0)	5.0-9.0
Primary progressive	35	28.6	47 (11.9)	23-73	26.0 (15.7)	10-31	8.0 (0.9)	5.5-9.0
Secondary progressive	166	31.3	44 (11.1)	22-73	20.5 (9.5)	10-100	7.0 (0.5)	5.0-9.0
Stroke	82	53.0	52 (16.9)	19-87	49.5 (35.1)	9-148	-	-

DOS=duration of stay.

effect size the greater the responsiveness of an instrument. By relating change scores to the variability of the study sample, effect sizes transform raw change scores with limited meaning to a standard metric thereby allowing comparison of different instruments and different samples. When instruments are compared in the same sample a direct indication of their relative responsiveness is provided. Under these circumstances the instrument with the largest effect size is considered the most responsive.³³ In addition, paired *t* tests were used to determine the statistical significance of disability change scores.

Results

SAMPLE

Table 1 presents the characteristics of the 283 patients studied. The multiple sclerosis group (71% of sample) contained more women, was slightly younger, and had a shorter length of stay than the stroke group (29% of sample). The EDSS scores indicated that the multiple sclerosis group were moderate to severely disabled.

APPROPRIATENESS

Table 2 presents BI and FIM score distributions for patients with multiple sclerosis and stroke patients on admission to the NRU. For

both disease groups, patient scores on the BI, FIM total, and FIM motor scales spanned the entire scale range, had mean scores near the scale midpoint, and had small floor and ceiling effects. These results indicate that these three scales are appropriate to the study samples. However, the results shown in table 2 raise concerns over the appropriateness of the FIM cognitive scale as a measure of cognitive disability in the patients with multiple sclerosis studied. Actual scores only span the upper (less disabled) range of the scale, the mean score is well above the scale midpoint, the SD is small, and the ceiling effect is only just below the recommended upper limit. The FIM cognitive scale is, however, more appropriate to the stroke than the multiple sclerosis sample.

RESPONSIVENESS

Table 3 presents disability change scores with their statistical significance and effect sizes for the BI and three FIM scales in patients with multiple sclerosis and stroke patients. Change scores for all scales in both disease groups were positive, indicating less disability on discharge than admission. These change scores were statistically significant ($p < 0.0001$) except for the FIM cognitive score in the multiple sclerosis group.

Effect sizes for the BI, FIM total, and FIM motor scales were very similar in each disease group indicating comparable responsiveness for these three scales. Also, in both disease groups effect sizes for the FIM cognitive scale were much less than for the BI, FIM total, and FIM motor scales indicating that the FIM cognitive scale is the least responsive scale.

Discussion

In this study the appropriateness and responsiveness of the BI and FIM were compared in patients with multiple sclerosis and stroke patients receiving inpatient neurorehabilitation. The results show that all measures were appropriate to the samples studied although the FIM cognitive scale has a notable ceiling effect in patients with multiple sclerosis. More importantly, the BI, FIM total, and FIM motor scales show similar responsiveness in both disease groups.

Appropriateness of disability measures, as defined in this study, is rarely reported in clinical studies. However, when scales measure a restricted range of health status it is important to show the appropriateness of this range to the study sample. The patients in this study had moderate to severe disability as measured by the EDSS. However, the range of cognitive dysfunction measured by the FIM cognitive

Table 2 BI and FIM scores on admission: sample range, mean, floor, and ceiling effect in two disease groups

Scale range	n	Sample range	Admission score mean (SD)	Floor effect n (%)	Ceiling effect n (%)
BI (0-20)					
MS	201	0-20	12.0 (5.7)	3 (1.5)	11 (5.5)
Stroke	82	0-20	11.4 (5.5)	1 (1.2)	7 (8.5)
FIM total (18-126)					
MS	201	24-122	89.4 (23.0)	0 (0)	0 (0)
Stroke	82	21-123	82.5 (26.8)	0 (0)	0 (0)
FIM motor (13-91)					
MS	201	13-88	59.1 (20.4)	4 (2.0)	0 (0)
Stroke	82	13-91	56.1 (20.9)	1 (1.2)	1 (1.2)
FIM cognitive (5-35)					
MS	201	11-35	30.3 (5.1)	0 (0)	36 (17.9)
Stroke	82	5-35	26.4 (7.9)	1 (1.2)	11 (13.4)

MS=multiple sclerosis; floor (ceiling) effects are the percentage of patients scoring the minimum (maximum) possible scores.

Table 3 Comparison of BI and FIM change scores, *p* values, and effect sizes

Disease groups	Change score mean (SD)	<i>p</i> Value	Effect size
MS patients			
BI	2.1 (2.4)	< 0.0001	0.37
FIM total	6.9 (8.3)	< 0.0001	0.30
FIM motor	6.9 (7.2)	< 0.0001	0.34
FIM cognitive	0.1 (2.9)	0.961 (NS)	0
Stroke patients			
BI	5.2 (4.4)	< 0.0001	0.95
FIM total	21.9 (19.0)	< 0.0001	0.82
FIM motor	19.1 (16.1)	< 0.0001	0.91
FIM cognitive	2.8 (4.8)	< 0.0001	0.61

Effect size=change score/standard deviation of admission score.

scale was restricted in these patients suggesting that this scale has limited usefulness for the measurement of cognitive disability in patients with multiple sclerosis undergoing neurorehabilitation. Even for the stroke patients the ceiling effect of the FIM cognitive scale is notable (13.4%) raising some concerns over its use in this patient group.

Patients in this study are not necessarily representative of multiple sclerosis or stroke patients undergoing neurorehabilitation. They were not randomly selected, and severely cognitively impaired patients were not represented as reasonable cognition was one of the selection criteria for admission to the unit. These considerations underlie the need to examine the appropriateness of scales to a study sample.

The most important finding of this study is the demonstration that the BI, FIM total, and FIM motor scales have similar responsiveness. This is perhaps surprising as the FIM was developed specifically to be more “sensitive to change” (responsive) than the BI,^{12 14 17} and has more items and more response categories. The findings of this study suggest that the FIM has no advantages over the BI in evaluating changes in disability due to therapeutic interventions. This has important clinical implications as the BI is quicker and simpler to rate. In addition, it can be rated by any healthcare professional whereas the developers of the FIM recommend rating by consensus opinion of a multidisciplinary team after a period (up to 72 hours) of patient observation. Furthermore, the BI can be administered by self report, adding to its impact on the design and cost of clinical studies.^{34 35}

Examining relative responsiveness is important as it helps clinicians to choose between competing disability measures on an empirical basis. The more responsive a disability measure, the more useful it is for evaluative studies as the importance of responsiveness lies in the trade off between sample size and statistical power.³⁶ For a given sample size, using a more responsive instrument increases the possibility of detecting a statistically significant result. Similarly, for a given statistical power a smaller sample size can be used if a more responsive instrument is employed.

There is no consensus as to which of the many methods of reporting responsiveness should be used. The effect size statistic used in this study is widely used and recommended.³⁷ However, different studies often use different statistical methods, thereby complicating comparative data interpretation. Furthermore, the responsiveness of instruments seems to be disease dependent. In this study effect sizes for BI and all FIM scales are greater for stroke patients than for patients with multiple sclerosis suggesting that these instruments are more responsive in stroke patients. Consequently, examining the responsiveness of competing instruments in the same samples undergoing the same interventions provides the best indication of relative responsiveness.

In conclusion, these results show that the BI, FIM total, and FIM motor scales have a simi-

lar ability to detect change in disability in a selected sample of multiple sclerosis and stroke patients undergoing neurorehabilitation. All measures were shown to be very appropriate to the study sample, although concerns are raised about using the FIM cognitive scale in patients with multiple sclerosis.

We thank medical, nursing and all therapy staff at the NRU for their involvement.

- 1 Sackett DL, Rosenberg WMC, Gray JAM, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.
- 2 Fleiss JL. *The design and analysis of clinical experiments*. New York: Wiley, 1986.
- 3 Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27-36.
- 4 Dodds TA, Martin DP, Stolow WC, *et al.* A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993;74:531-6.
- 5 Carey RG, Seibert JH, Posavac EJ. Who makes the most progress in inpatient rehabilitation? An analysis of functional gain. *Arch Phys Med Rehabil* 1988;69:337-43.
- 6 Kidd D, Howard RS, Losseff NA, *et al.* The benefit of inpatient neurorehabilitation in multiple sclerosis. *Clin Rehabil* 1995;9:198-203.
- 7 Greenspun BG, Stineman M, Agri R. Multiple sclerosis and rehabilitation outcome. *Arch Phys Med Rehabil* 1987;68:434-7.
- 8 Aisen ML, Sevilla D, Fox N. Inpatient rehabilitation for multiple sclerosis. *J Neurol Rehabil* 1996;10:43-6.
- 9 Freeman JA, Langdon DW, Hobart JC, *et al.* The impact of inpatient rehabilitation on progressive multiple sclerosis. *Ann Neurol* 1997;42:236-44.
- 10 Gompertz P, Pound P, Ebrahim S. The reliability of stroke outcome measures. *Clin Rehabil* 1993;7:290-6.
- 11 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27(suppl):S217-32.
- 12 Keith RA, Granger CV, Hamilton BB, *et al.* The functional independence measure: a new tool for rehabilitation. In: Eisenberg M, Grzesiak R, eds. *Advances in clinical rehabilitation*. New York: Springer Verlag, 1987:6-18.
- 13 Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Maryland State Medical Journal* 1965;14:61-5.
- 14 Granger C, Hamilton B, Keith R, *et al.* Advances in functional assessment for medical rehabilitation. *Topics in Geriatric Rehabilitation* 1986;1:59-74.
- 15 Wade DT, Collin C. The Barthel ADL index: a standard measure of disability? *International Disability Studies* 1988;10:64-7.
- 16 Collin C, Wade DT, Davis S, *et al.* The Barthel ADL index: a reliability study. *International Disability Studies* 1988;10:61-3.
- 17 Hamilton B, Granger C, Sherwin F, *et al.* A uniform national data system for medical rehabilitation. In: Fuhrer MJ, ed. *Rehabilitation outcomes: analysis and measurement*. Baltimore: Brookes, 1987:137-47.
- 18 Kidd D, Stewart G, Baldry J, *et al.* The functional independence measure: a comparative validity and reliability study. *Disabil Rehabil* 1995;17:10-14.
- 19 Hobart JC, Lamping DL, Freeman JA, *et al.* Measuring disability in multiple sclerosis: reliability of the functional independence measure [abstract]. *J Neurol* 1996;243(6 suppl 2):S32.
- 20 Hamilton BB, Laughlin JA, Fielder RC, *et al.* Interrater reliability of the 7-level functional independence measure (FIM). *Scand J Rehabil Med* 1994;26:115-19.
- 21 Granger CV, Cotter ACR, Hamilton BB, *et al.* Functional assessment scales: a study of persons with multiple sclerosis. *Arch Phys Med Rehabil* 1990;71:870-5.
- 22 Freeman JA, Playford ED, Nicholas RS, *et al.* A neurological rehabilitation unit: audit of activity and outcome. *J R Coll Physicians Lond* 1996;30:21-6.
- 23 Hobart J, Lamping D, Freeman J, *et al.* Measuring neurology—is bigger better? Comparative measurement properties of the functional independence measure (FIM) and the Barthel index (BI) [abstract]. *Neurology* 1997;48(3 suppl):A235.
- 24 Granger CV, Cotter AC, Hamilton BB, *et al.* Functional assessment scales: a study of persons after stroke. *Arch Phys Med Rehabil* 1993;74:133-8.
- 25 Hobart J, Langdon D, Lamping D, *et al.* Can cognitive disability in multiple sclerosis be measured from behavioural observation? Validity of the functional independence measure cognitive scale (FIM-c) [abstract]. *Multiple Sclerosis* 1997;3:268.
- 26 Kurtzke J. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444-52.
- 27 Willoughby EW, Paty DW. Scales for rating impairment in multiple sclerosis: a critique. *Neurology* 1988;38:1793-8.
- 28 Willoughby E. Impairment in multiple sclerosis. *Multiple Sclerosis Management* 1995;2:13-16.
- 29 Sharrack B, Hughes RAC. Clinical scales for multiple sclerosis. *J Neurol Sci* 1996;135:1-9.

- 30 Hobart JC, Lamping DL, Freeman JA, et al. Reliability, validity, and responsiveness of the Kurtzke expanded disability status scale (EDSS) in multiple sclerosis [abstract]. *J Neurol Neurosurg Psychiatry* 1997;62:212.
- 31 Holmes W, Shea J. Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. *Qual Life Res* 1997;6:561-71.
- 32 Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;40:171-8.
- 33 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(3 suppl):S178-89.
- 34 Gompertz P, Pound P, Ebrahim S. A postal version of the Barthel index. *Clin Rehabil* 1994;8:233-9.
- 35 Hobart JC, Lamping DL, Thompson AJ. Measuring disability in neurological disease: validity of the self-report Barthel index [abstract]. *J Neurol* 1996;243(suppl 2):S25.
- 36 Liang MH, Larson MG, Cullen KE, et al. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985;28:542-7.
- 37 Scientific advisory committee of the Medical Outcomes Trust. Instrument review criteria. *Medical Outcomes Trust Bulletin* 1995;3:1-IV.