Monitoring recovery by physical therapists using the FIM scale during rehabilitation programs: An inter-rater and intra-rater reproducibility study[*]

Tommasina Russo[1], Giorgio Felzani[2], Mario Giunta[1], Cristina Di Mascio[2], Carmine Marini[1#]

[1]Department of Medicina Interna e Sanità Pubblica, University of L'Aquila, L'Aquila, Italy;
[#]Corresponding Author: marini@cc.univaq.it, nina.russo@cc.univaq.it, mariogiunta1@libero.it
[2]Casa di Cura "San Raffaele" di Sulmona, L'Aquila, Italy; giorgio.felzani@sanraffaele.it, cris.dimascio@gmail.com

ABSTRACT

Our aim was to evaluate the reproducibility of the Functional Independence Measure (FIM) scale when assessed by physical therapists in the routine setting of a Rehabilitation Hospital. We included a consecutive series of patients with spinal cord or cerebral lesions. Each of the 50 selected patients was evaluated by two of the 5 experienced physical therapists participating in the study. The degree of inter-rater and intra-rater agreement was measured by a weighted k statistic, k for perfect agreement, and k for the agreement with tolerance. The weighted k index for inter-rater agreement on the FIM score was in the almost perfect range (k 0.87; 95% CI = 0.79 - 0.95), but a 20-point tolerance was necessary to reach a k value of 0.81 (95% CI = 0.66 - 0.95). Agreement was substantial or almost perfect for most subscales, but the k index with 1-point tolerance reached the almost perfect rating for comprehension only. For intra-rater agreement, weighted k index was in the almost perfect range for the FIM score and for all subscales; kappa index reached the almost perfect range with a 4-point tolerance for FIM score and with 1-point of tolerance for all subscales except interpersonal relations. FIM is useful to monitor patient improvement during rehabilitation treatment, mostly when assessed by the same physical therapist.

Keywords: Reproducibility; Rehabilitation;

Functional Independence Measure

1. INTRODUCTION

Rehabilitation services require instruments that are suitable for the following patients over time customizing rehabilitation protocols for measuring disability. Moreover, disability rating scales that may be assessed by physical therapists would be of greatest interest favouring a closer monitoring of clinical course and lower costs [1,2]. The FIM is considered a useful support for clinical practice in each rehabilitation area, although this scale seems less useful for spinal cord injury [1-3]. It is an easy-to-use, standardised and robust general measure of functional disability. Previous studies showed high intra-rater and inter-rater reliability of the FIM, indicating the internal consistency of the scale although it was not sensitive enough to assess changes in patients with tetraplegia [1,4-11]. However, these studies failed to clarify what was the minimum change that may reflect a real change of patients' status rather than random variability. Actually, to be helpful in monitoring recovery during the treatment of a scale should be suitable for physical therapists and should prove reliable enough to show even minimal changes in the patient's disability.

Reproducibility studies estimate the probability that the same score is attributed when the patient is retested and then the likelihood that an improved score reflects a true clinical improvement. Besides, reproducibility also reflects the reliability of the scale and training of raters. Scales showing a high reproducibility index when administrated by physical therapists may be adopted for monitoring the clinical course of patients during rehabilitation programs. Therefore, we examined the repro-

ducibility of the FIM scale when assessed by physical therapists in a sample of patients with mild to severe disability in the routine setting of a Spinal Unit and a Department of Physical Medicine of a Rehabilitation Hospital.

2. METHODS

In the present study we included a consecutive case series of 50 patients admitted to the Spinal Unit and the department of Physical medicine of the Rehabilitation hospital "Casa di Cura San Raffaele" of Sulmona, L'Aquila, Italy, between March and August 2009, because of the occurrence of neurological deficits caused by spinal cord injury, neurodegenerative, vascular or inflammatory diseases. All patients provided in person informed consent, according to national and international regulations. Comatose patients were excluded. FIM is an 18-subscale ordinal scale which rates the level of assistance required to perform various activities of daily living using a seven-level scoring system, with scores ranging between 126 (normal status) and 18 (totally dependent) [12]. Five experienced physical therapists, that routinely used the FIM scale, were arranged in 25 combinations in a balanced design in which each physical therapist was in turn once the first and once the second rater. Each couple of physical therapists evaluated 2 patients, assigned the score to each of the 18 subscales and computed the total score. Each patient was randomly assigned to one of the 25 couples of raters and was evaluated twice, within an interval of 24 hours (+/−5 hours). The raters were taught to independently evaluate each patient and not to communicate the scores to each other or to the patient in order to keep independency of the assessments. In ten instances the first and the second raters corresponded to the same physical therapist. These patients were used to evaluate intra-rater reliability, while the remaining 40 patients were used for inter-rater reliability.

We performed a graphical descriptive analysis of the FIM scores in order to evaluate the distribution of discrepancies over the scale range. The degree of inter-rater and intra-rater agreement was measured with weighted k statistic, accounting for severity of disagreement. Moreover, k for perfect agreement, and k for the agreement with tolerance were also computed in order to evaluate the minimum variation reflecting a true change in patient status [13]. The indexes were computed for each couple of raters and then an overall k index was calculated according to the method proposed by Fleiss *et al*. [14] The values of the k statistic were interpreted according o the criteria of Landis and Koch [15]. For a k index < 0.00 agreement was termed as poor; for a k index between 0.00 and 0.20, as slight; for a k index between 0.21 and 0.40, as fair; for a k index between 0.41 and 0.60, as

moderate; for a k index between 0.61 and 0.80, as substantial; for a k index between 0.81 and 1.00, as almost perfect [15].

3. RESULTS

Our study population included 50 patients (28 men and 22 women) with spinal cord (50%) or cerebral (50%) lesions referred to the Rehabilitation Center of "Casa di Cura San Raffaele" of Sulmona, Italy: 48 were first choices while 2 were replacement choices due to death or to unexpected discharge before completion of the protocol. The mean age was 59.5 +/− 22.58 years. Etiology and distribution of neurological deficits for the included patients were reported in **Table 1**. Graphical analysis showed a uniform distribution of the sample over the range of the FIM scale (**Figures 1(a)** and **(b)**) and of each subscale with the exceptions of *personal care*, *feeding oneself*, *sphincter al control*, *communication*, *relational/ cognitive capacity* in which more values occurred in the higher range of the scale and of *locomotion*, in which most values occurred in the lower range. The maximum disagreement on the FIM scale produced a 40-point difference between raters. Outlier values of disagreement were observed for a few patients in several subscales, with differences of more than 3 points in 12 of 18 sub-

Table 1. Main characteristics of the study population.

Parameters	N	%
Patients	50	100
Males	28	56.0
Age (mean ± SD)	59.46 ± 22.58	
Diagnosis		
Spinal cord injury	15	30.0
Vascular disease	13	26.0
Spondylosis and disc herniation	7	14.0
Parkinsonism	4	8.0
Myelitis	2	4.0
Cerebral palsy	2	4.0
Diabetic peripheral neuropathy	1	2.0
Multiple sclerosis	1	2.0
Myopathy	1	2.0
Arnold-Chiari disease	1	2.0
Miscellanea	3	6.0
Neurological deficits		
Paraparesis/paraplegia	16	32.0
Tetraparesia/tetraplegia	12	24.0
Hemiparesis/hemiplegia	9	18.0
Extrapiramidal signs	4	8.0
Peripheral nerve deficits	1	2.0
Other neurological deficits	2	4.0
Deficit in multiple functional systems	6	12.0

$K = 0.88; 95\%\ CL: 0.79 - 0.95$

(a)

$K = 0.98; 95\%\ CL: 0.96 - 1.00$

(b)
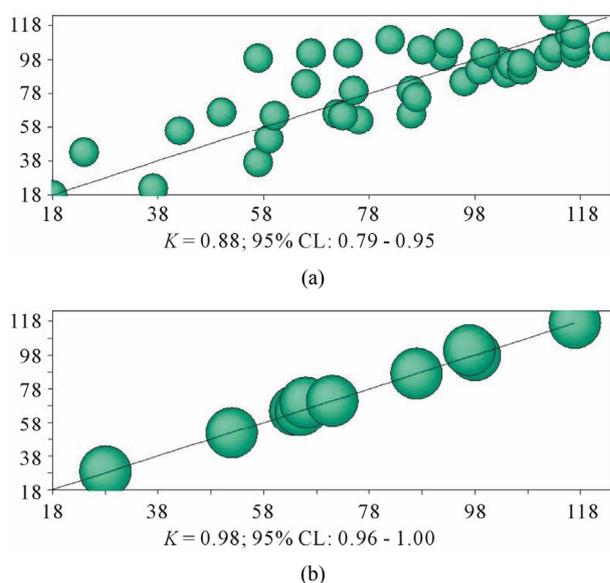
Figure 1. Distribution of FIM overall scores of intra-rater agreement analysis.

scales (tyding oneself, washing oneself, dressing from the waist up, dressing from the waist down, perineal hygiene, bladder control, bowel control, water closet, walking/wheelchair, stairs, interpersonal relations, and problem solving). Kappa indexes and 95% CI for inter-rater agreement of FIM overall score and for subitems are reported in **Table 2**. Based on weighted *k* index, the agreement on the overall FIM score was almost perfect (*k* 0.87; 95% CI = 0.79 - 0.95). The agreement was substantial or almost perfect in all scales but that relating to *walking/wheelchair* in which agreement was moderate. The *k* index of perfect agreement for the FIM overall score was slight (*k* 0.18; 95% CI = 0.006 - 0.30). A 20-point tolerance was necessary to reach a *k* value rated as almost perfect (*k* 0.81; 95% CI = 0.66 - 0.95). The kappa index of perfect agreement was fair for all subscales of FIM with the exception of bowel control, stairs, and problem solving for which it was moderate. The agreement with 1-point of tolerance reached the almost perfect rating for comprehension only (*k* 0.82; 95% CI = 0.64 - 1.00) and was substantial for the majority of the remaining subscales but tyding oneself, washing oneself, dressing from the waist up, dressing from the waist down, walking/wheelchair, and interpersonal relations in which it was moderate. Kappa indexes and 95% CI of intra-rater agreement for FIM overall score and for subscales (**Table 3**) were always higher than the corresponding values of inter-rater agreement. The weighted *k* index was in the almost perfect range for the overall FIM score and for all subscales. The kappa index of perfect agreement, for the FIM overall score was substantial (*k* 0.77; 95% CI = 0.48 - 1.00) and became almost perfect with a 4-point tolerance. The analysis of intra-rater perfect

agreement showed *k* values rated as almost perfect or substantial for all subscales but dressing from the waist down in which it was moderate. The agreement with 1-point of tolerance reached the almost perfect level for all subscales.

4. DISCUSSION

The results of this study including a sample of patients with a mild to severe level of disability indicate that inter-rater and intra-rater reproducibility of FIM are high when evaluated by the weighted *k* index accounting for severity of disagreement. However, the analysis of perfect agreement and of agreement with tolerance indicated that a 20-point tolerance was necessary to reach a substantial inter-rater agreement while with a 4-point tolerance intra-rater agreement was almost perfect. Kappa index with 1-point tolerance showed a substantial or almost perfect inter-rater agreement in most subscales and almost perfect intra-rater agreement in all subscales. Reproducibility of scales may be influenced by the distribution of values across the scale range, being higher when values fall in a tighter range. In the present study, the overall FIM and subscales scores are spread across the whole range in most of the scales. Therefore, the sample is fairly representative of cases seen in common clinical practice and allowed an unbiased estimation of reproducibility. The time interval of 24 hours between the first and the second observations may have favoured intra-rater agreement that was almost perfect in almost all subscales. However, had we adopted a wider time interval among assessments, changes in the functional status of patients might have produced a spurious disagreement. We think that a 24-hour interval was an acceptable compromise. Moreover, the good agreement might have also depended on the inclusion of patients in the post-acute phase in a well-defined setting, with examiners trained in the identification of relevant clinical features, actively collaborating and exchanging views on patients course. On the other hand, routine administration of FIM might have reduced the agreement in the long run, due to the tendency to over-interpret some subscales. Previous studies indicate that FIM provides good inter- and intra-rater reliability across a wide variety of raters with different professional backgrounds and levels of training, but few studies addressed the reproducibility of FIM in patients with spinal cord injury since it was considered not specifically designed for those subjects [1,4,5,16,17]. We confirmed the high reproducibility of the scale in a sample including 50% of subjects with spinal cord lesions. An important lesson from our study and literature reports is that the level of reproducibility of FIM is high even when the scale is assessed by physical therapists, although one must consider as a source of variability the level of professional skill, experience with the FIM and

Table 2. Kappa indexes for inter-rater reproducibility.

	Weighted		Perfect agreement		One-point tolerance	
	k	95% CI	k	95% CI	k	95% CI
Personal care						
Feeding oneself	0.80	0.68 - 0.92	0.28	0.09 - 0.47	0.74	0.56 - 0.93
Tidying oneself	0.70	0.54 - 0.85	0.22	0.05 - 0.38	0.59	0.37 - 0.80
Washing oneself	0.69	0.54 - 0.84	0.17	0.01 - 0.34	0.52	0.30 - 0.73
Dressing, flrom the waist up	0.71	0.56 - 0.86	0.27	0.11 - 0.43	0.60	0.41 - 0.80
Dressing, from the waist down	0.75	0.62 - 0.88	0.21	0.06 - 0.37	0.51	0.29 - 0.73
Perineal hygiene	0.76	0.63 - 0.89	0.27	0.11 - 0.42	0.67	0.48 - 0.85
Sphincteral control						
Bladder control	0.86	0.76 - 0.95	0.40	0.24 - 0.57	0.80	0.65 - 0.95
Bowel control	0.78	0.65 - 0.91	0.42	0.26 - 0.59	0.63	0.45 - 0.82
Mobility (Transfers)						
Bed-chair-wheelchair	0.87	0.79 - 0.94	0.36	0.20 - 0.52	0.80	0.65 - 0.95
Water closet	0.83	0.72 - 0.94	0.40	0.24 - 0.56	0.74	0.57 - 0.90
Bath or shower	0.80	0.70 - 0.91	0.37	0.21 - 0.53	0.69	0.51 - 0.88
Locomotion						
Walking, wheelchair	0.60	0.40 - 0.80	0.29	0.11 - 0.47	0.41	0.15 - 0.66
Stairs	0.87	0.79-0.96	0.43	0.28 - 0.57	0.80	0.65 - 0.96
Communication						
Comprehension	0.84	0.71 - 0.97	0.35	0.16 - 0.54	0.82	0.64 - 1.00
Expression	0.79	0.62 - 0.96	0.36	0.16 - 0.56	0.74	0.50 - 0.98
Relational/cognitive capacity						
Interpersonal relations	0.73	0.54 - 0.92	0.28	0.10 - 0.47	0.59	0.35 - 0.84
Problem solving	0.83	0.68 - 0.98	0.51	0.34 - 0.68	0.74	0.53 - 0.95
Memory	0.80	0.67 - 0.94	0.33	0.14 - 0.52	0.67	0.45 - 0.89
FIM overall	0.87	0.79 - 0.95	0.18	0.06 - 0.30	0.29	0.13 - 0.44

acquaintance with the patient [1,2,5]. According to our results the FIM scale may be adequately assessed by treating physical therapists when adequately trained to assess the scale in the routine clinical practice, with uniform reproducibility across all subscales. A sound difference between the levels of inter- and intra-rater agreement is evident from our study. A possible explanation of this result may consist in the misinterpretation of scale coding by some raters and might indicate the need of further training or loss of adherence to coding rules with time. Therefore, periodic retraining of raters should be planned in order to keep high the reproducibility. Several subscales did not reach a kappa index in the almost perfect or substantial range when inter-rater agreement was assessed with 1-point of tolerance, while all subscales reached the almost perfect rating of intra-rater agreement with the same analysis. So if the evaluations are carried out by the same rater, changes as small as 1-point in each subscale should be considered clinically significant. Otherwise, when evaluations are performed by different raters, variations of less than 2 points in each sub-scale might be a consequence of variability in the assessments. Our results thus suggest that the repeated administration of FIM by the treating physical therapist may record even small variations during the rehabilitation program. This practice may produce a reinforcing effect in terms of engagement of the patient with therapy. Moreover, precise monitoring of functional status allows a continuous adaptation of the rehabilitation protocol favouring the achievement of the best improvement in the patient's physical performance and the least rate of complications. However, monitoring of patients with the FIM scale should be performed by the same physical therapist to minimize random variability of assessments.

Table 3. Kappa indexes for intra-rater reproducibility.

	Weighted		Perfect agreement		One-point tolerance	
	k	95% CI	k	95% CI	k	95% CI
Personal care						
Feeding oneself	0.99	0.96 - 1.00	0.88	0.65 - 1.00	1.00	-
Tidying oneself	0.96	0.90 - 1.00	0.74	0.44 - 1.00	1.00	-
Washing oneself	0.97	0.93 - 1.00	0.75	0.46 - 1.00	1.00	-
Dressing, from the waist up	0.99	0.96 - 1.00	0.88	0.65 - 1.00	1.00	-
Dressing, from the waist down	0.96	0.93 - 0.99	0.64	0.31 - 0.96	1.00	-
Perineal hygiene	0.99	0.97 - 1.00	0.87	0.64 - 1.00	1.00	-
Sphincteral control						
Bladder control	1.00	1.00 - 1.00	1.00	-	1.00	-
Bowel control	1.00	1.00 - 1.00	1.00	-	1.00	-
Mobility (Transfers)						
Bed-chair-wheelchair	0.99	0.97 - 1.00	0.88	0.67 - 1.00	1.00	-
Water closet	1.00	1.00 - 1.00	1.00	-	1.00	-
Bath or shower	1.00	1.00 - 1.00	1.00	-	1.00	-
Locomotion						
Walking, wheelchair	0.98	0.95 - 1.00	0.86	0.60 - 1.00	1.00	-
Stairs	0.99	0.97 - 1.00	0.79	0.46 - 1.00	1.00	-
Communication						
Comprehension	0.98	0.94 - 1.00	0.85	0.60 - 1.00	1.00	-
Expression	1.00	1.00 - 1.00	1.00	-	1.00	-
Relational/cognitive capacity						
Interpersonal relations	0.89	0.75 - 1.00	0.71	0.40 - 1.00	1.00	-
Problem solving	0.99	0.95 - 1.00	0.86	0.62 - 1.00	1.00	-
Memory	0.97	0.92 - 1.00	0.73	0.42 - 1.00	1.00	-
FIM overall	0.98	0.96 - 1.00	0.77	0.48 - 1.00	0.77	0.48 - 1.00

5. CONCLUSION

In conclusion, according to our study, FIM when assessed by physical therapists may be very useful in the management of patients in a rehabilitation setting. However, to achieve a high level of agreement, the scale should be administered by the same rater and regular retraining courses should be recommended.

6. ACKNOWLEDGEMENTS

REFERENCES

[1] Anderson, K., Aito, S., Atkins, M., Biering-Sørensen, F., Charlifue, S., Curt, A., Ditunno, J., Glass, C., Marino, R., Marshall, R., Mulcahey, M.J., Post, M., Savic, G., Scivoletto, G. and Catz, A. (2008) Functional recovery measures for spinal cord injury: An evidence-based review for clinical practice and research. *Journal of Spinal Cord Medicine*, **31**, 133-144.

[2] Tischler, H., Platzer, A., Vian, P. and Genetti, B. (2001) The FIM scale as a planning tool for medical, nursing and physiotherapy requirements in rehabilitation. Use in a recovery and functional rehabilitation unit at Merano Hospital. *European Journal of Physical and Rehabilitation Medicine*, **37**, 39-50.

[3] Dodds, T.A., Martin, D.P., Stolov, W.C. and Deyo, R.A. (1993) A validation of the functional independence meas-

urement and its performance among rehabilitation inpatients. *Archives of Physical Medicine and Rehabilitation*, **74**, 531-536. doi:10.1016/0003-9993(93)90119-U

[4] Segal, M.E., Ditunno, J.F. and Staas, W.E. (1993) Interinstitutional agreement of individual functional independence measure (FIM) items measured at two sites on one sample of SCI patients. *Paraplegia*, **31**, 622-631. doi:10.1038/sc.1993.101

[5] Ottenbacher, K.J., Hsu, Y., Granger, C.V. and Fieldler, R.C. (1996) The reliability of the functional independence measure: A quantitative review. *Archives of Physical Medicine and Rehabilitation*, **77**, 1226-1231. doi:10.1016/S0003-9993(96)90184-7

[6] Kidd, D., Stewart, G., Baldry, J., Johnson, J., Rossiter, D., Petruckevitch, A. and Thompson, A.J. (1995) The Funtional Independence Measure: A comparative validity and reliability study. *Disability and Rehabilitation*, **17**, 10-14. doi:10.3109/09638289509166622

[7] Hamilton, B.B., Laughlin, J.A., Fiedler, R.C. and Granger, C.V. (1994) Inter-rater reliability of the seven level Functional Independence Measure (FIM). *Archives of Physical Medicine and Rehabilitation*, **26**, 115-119.

[8] Chau, N., Daler, S., Andrew, J. and Patris, A. (1994) Inter-rater agreement of two functional independence scales the Functional Independence Measure (FIM) and a subjective uniform continuous scale. *Disability and Rehabilitation*, **16**, 63-71. doi:10.3109/09638289409166014

[9] Ottenbacher, K.J., Mann, W.C., Granger, C.V., Tomita, M., Hurren, D., Charvat, B. (1994) Inter-rater agreement and stability of functional assessment in the community-based elderly. *Archives of Physical Medicine and Rehabilitation*, **75**, 1297-1301.

[10] Fricke, J., Unsworth, C. and Worrell, D. (1993) Reliability of the functional independence measure with occupational therapists. *Australian Occupational Therapy Journal*, **40**, 7-15. doi:10.1111/j.1440-1630.1993.tb01770.x

[11] Daving, Y., Andren, E., Nordholm, L. and Grimby, G. (2001) Reliability of an interview approach to the functional independence measure. *Clinical Rehabilitation*, **15**, 301-310. doi:10.1191/026921501669986659

[12] Keith, R.A., Granger, C.V., Hamilton, B.B. and Sherwin, F.S. (1987) The functional independence measure: A new tool for rehabilition. *Advances in Clinical Rehabilitation*, **1**, 6-18.

[13] Kramer, M.S. and Feinstein, A.R. (1983) Clinical biostatistics, LIV: The biostatistics of concordance. *Clinical Pharmacology & Therapeutics*, **29**, 111-123. doi:10.1038/clpt.1981.18

[14] Fleiss, J.L. (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin*, **76**, 378-382. doi:10.1037/h0031619

[15] Landis, J.R. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159-174. doi:10.2307/2529310

[16] Gabbe, B.J., Sutherland, A.M., Wolf, R., Williamson, O.D. and Cameron, P.A. (2007) Can the modified functional independence measure be reliably obtained from the patient medical record by different raters? *The Journal of Trauma*, **63**, 1374-1379. doi:10.1097/01.ta.0000240481.55341.38

[17] Dallmeijer, A.J., Dekker, J., Roorda, L.D., Knol, D.L., van Baalen, B., de Groot, V., Schepers, V. and Lankhorst, G.J. (2005) Differential item functioning of the functional independence measure in higher performing neurological patients. *Journal of Rehabilitation Medicine*, **37**, 346-352. doi:10.1080/16501970510038284