Scand J Rehab Med 26: 115–119, 1994

INTERRATER RELIABILITY OF THE 7-LEVEL FUNCTIONAL INDEPENDENCE MEASURE (FIM)

Byron B. Hamilton, MD, PhD, Judith A. Laughlin, RN, PhD, Roger C. Fiedler, PhD, and Carl V. Granger, MD

From the Center for Functional Assessment Research, Department of Rehabilitation Medicine, School of Medicine and Biomedical Sciences, State University of New York, Buffalo, NY, USA

ABSTRACT. The Functional Independence Measure (FIM) is an 18-item, 7-level scale developed to uniformly assess severity of patient disability and medical rehabilitation functional outcome. FIM interrater reliability in the clinical setting is reported here. Clinicians from 89 US inpatient comprehensive medical rehabilitation facilities newly subscribing to the uniform Data System for Medical Rehabilitation from January 1988-June 1990 evaluated 1018 patients with the FIM. FIM total, domain and subscale score intraclass correlation coefficients (ICC) were calculated using ANOVA; FIM item score agreement was assessed with unweighted Kappa coefficient. Total FIM ICC was 0.96; motor domain 0.96 and cognitive domain 0.91; subscale score range: 0.89 (social cognition) to 0.94 (self-care). FIM item Kappa range: 0.53 (memory) to 0.66 (stair climbing). A subset of 24 facilities meeting UDSMR data aggregation reliability criteria had Intraclass and Kappa coefficients exceeding those for all facilities. It is concluded that the 7-level FIM is reliable when used by trained/tested inpatient medical rehabilitation clinicians.

Key words: rehabilitation, disability evaluation, test reliability.

The primary clinical objective of comprehensive inpatient medical rehabilitation is to reduce patient disability (14)[1] by increasing independence in performance of activities of daily living. An instrument designed to assess person level of disability in this setting is the Functional Independence Measure (FIM), developed by a joint task force of the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation (8, 12, 15). The FIM was intended to be used as a uniform measure of severity of disability and rehabilitation functional outcome and was designed, evaluated and modified in three phases (12).

The FIM scale consists of 18 items each assessed on 7 levels which, when summed, may be used to estimate a person's need for assistance (burden of care) or resource cost of disability (9, 10, 12). In addition to a total score, the FIM provides two domain scores (motor and cognitive), six subscale scores (self-care, sphincter control, transfers, locomotion, communication and social cognition), and 18 individual item scores. The items are listed in Table I and generic scale levels are summarized in Table II. Each FIM item has a more specific set of scale level descriptors than appear in the generic scale in Table II (See reference 11 for details). An original 4-level FIM scale was increased to 7 levels in 1987 on the recommendation of clinicians, in order to increase sensitivity (11).

This report presents the results of a study of FIM interrater reliability among clinicians in inpatient comprehensive medical rehabilitation facilities subscribing to the Uniform Data System for Medical Rehabilitation (UDSMR). The UDSMR provides a Data Management Service for medical rehabilitation facilities and includes the FIM as the functional assessment component of the data set. These data are used by facilities to determine severity of disability of patients on admission, to measure functional gain, to estimate efficiency and compare outcomes with facilities in their region and nationality.

The FIM is also utilized by inpatient medical rehabilitation facilities in Australia, Canada, France, Japan, Italy, Germany, Portugal, and Sweden.

METHODS

The sample for this study included all 89 US freestanding

[1]The terms 'impairment and disability' referred to in this report are as defined by the World Health Organization (14).

are the subject of this report. Preliminary results of this report have been published in abstract form (13). The interrater reliability of the earlier (1984–86) development phase 4-level FIM has been reported previously (12). Further, internal consistency of the 7-level FIM has been reported to be high (0.93), and sensitivity to change significant (4).

The results reported here indicated that the interrater reliability of the 7-level FIM was acceptably high, both for all first-time respondent facility clinicians and particularly for those meeting UDSMR data aggregation reliability criteria. These latter criteria have been used to select which facilitiy data would be aggregated into the regional and national data reports.

Data from facilities not meeting these criteria were reported back to the facilities, but were not aggregated into regional and national reports. Facilities not meeting the criteria were given subsequent opportunities to do so and usually succeeded after one or two more trials.

For most clinicians mastery of functional assessment can probably not be achieved by only reading a training guide and/or viewing a training videotape. In order to achieve a high level of reliability appropriate training and testing are necessary. This is supported by Fricke et al. (7), who observed that FIM interrater reliability was highest for FIM-trained but previously FIM-inexperienced therapists assessing eight FIM items germane to occupational therapists. The implication of this is that functional assessment training cannot be casual; rather, it requires mastery.

The conventional method of interrater reliability, in this case assessing 10 or more patients by two or more clinicians in the rehabilitation facility's setting, demonstrated high interrater reliability of the FIM as a tool. In order to ensure that clinicians using the FIM in subscribing facilities are knowledgeable, in 1990, the UDSMR implemented a less cumbersome and more efficient method for credentialing. Credentialing is accomplished by testing clinician mastery of FIM definitions and application based on standardized written cases. Standardized tests have the advantages of controlling for a variety of impairments and severity of disability that influence difficulty of functional assessment (7), reducing scoring errors, and providing for efficiency and uniformity when testing a large number of facilities and clinicians in the United States and in other countries. In the future, written tests might be replaced with standardized videotaped cases in order to simulate the clinical behavior of patients more clearly.

It is concluded from the field testing approach for assessing interrater reliability reported here that the FIM 7-level scale has demonstrated high interrater reliability when used by clinicians meeting UDSMR criterion standards for comprehensive inpatient medical rehabilitation facilities in the United States. Further, mastery training and testing in functional assessment seem necessary. A medical rehabilitation data system must achieve high clinical sensitivity and reliability in order to provide comparability of patients and patient outcomes. Once achieved and broadly applied data from such a system will advance the scientific basis of medical rehabilitation practice and research.

ACKNOWLEDGEMENTS

The authors wish to thank Rich Kayton for programming and data processing, James Shelton, Carl Wende, Diane Cookfair and Maria Zielezny for preliminary statistical consultation and Kenneth Ottenbacher for final manuscript review.

This research was supported by funds from the Center for Functional Assessment Research, State University of New York at Buffalo.

REFERENCES

1. Armstrong, G. D.: The intraclass correlation as a measure of interrater reliability of subjective judgments. Nurs Res *30*: 314–315, 320A, 1981.
2. Bartko, J. J. & Carpenter, W. T.: On the methods and theory of reliability. J Nerv Mentl Dis *163*: 307–317, 1976
3. Cohen, J. A.: Coefficient of agreement for nominal scales. Educ Psychol Meas *20*: 37–46, 1960.
4. Dodds, T. A.: Martin, D. P., Stolov, W. C. & Deyo, R. A.: A validation of the Functional Independence Measurement and its performance among rehabilitation inpatients. Arch Phys Med Rehabil *74*: 531–536, 1993.
5. Fleiss, J. L.: Measuring nominal scale agreement among many raters. Psychol Bull *76*: 378–382, 1971.
6. Fleiss, J.: The measurement of interrater agreement. In: Statistical methods for rates and proportions. Wiley, New York, 1981, 218.
7. Fricke, J., Unsworth, C. & Worrell, D.: Reliability of the Functional Independence Measure with occupational therapists. Aust Occup Ther J *40*: 7–15, 1993.
8. Granger, C. V., Hamilton, B. B., Keith, R. A., Zielezny, M. & Sherwin, E. S.: Advances in functional assessment for medical rehabilitation. Top Geriatr Rehabil *1*: 59–74, 1986.
9. Granger, C. V., Cotter, A. C., Hamilton, B. B., Fielder, R. C. & Hens, M. M.: Functional assessment scales: a study of persons with multiple sclerosis. Arch Phys Med Rehabil *71*: 870–875, 1990.
10. Granger, C. V., Cotter, A. C., Hamilton, B. B. & Fiedler,

R. C.: Functional assessment scales: a study of persons after stroke. Arch Phys Med Rehabil 74: 133–138, 1993.
11. Guide for use of the uniform data set for medical rehabilitation. Buffalo: State University of New York at Buffalo, 1987. (Note: updated versions of the Guide were also published in 1990 and 1993.)
12. Hamilton, B. B., Granger, C. V., Sherwin, F. S., Zielezny, M. & Tashman, J. S.: A uniform national data system for medical rehabilitation. In: Fuhrer, M., (ed.). Rehabilitation outcomes: analysis and measurement. Brookes, Baltimore, 1987, 137–147.
13. Hamilton, B. B., Laughlin, J. A., Granger, C. V. & Kayton, R. M.: Interrater agreement of the seven-level Functional Independence Measure (FIM). Arch Phys Med Rehabil *72*: 790, 1991 (abstract).
14. International Classification of Impairments, Disabilities and Handicaps. World Health Organization, Geneva, 1980.
15. Keith, R. A., Granger, C. V., Hamilton, B. B. & Sherwin, F. S.: The Functional Independence Measure: a new tool for rehabilitation. In: Eisenberg, M. G. & Grzesiak, R. C. (eds.). Advances in Clinical Rehabilitation: Vol 1. Springer-Verlag, New York, 1987, 6–18.
16. Sheikh, K.: Disability scales: assessment of reliability. Arch Phys Med Rehabil *67*: 245–249, 1986.
17. Shrout, P. E. & Fleiss, J. L.: Intraclass correlations: uses in assessing rater reliability. Phychol Bull 86: 420–428, 1979.
18. SPSS-X user's guide, 2nd edition. SPSS, Inc. Chicago, 1986.
19. Tinsley, H. E. & Weiss, D. J.: Interrater reliability and agreement of subjective judgments. Journal of Counseling Psychology *22*: 358–376, 1975.
20. Winer, B. J.: Statistical Principles in Experimental Design. 2nd ed. McGraw-Hill, New York, 1971.

Address for offprints:

Carl V. Granger, MD
232 Parker Hall
SUNY South Campus
Buffalo, 14214 NY, USA