# RELIABILITY OF THE GROSS MOTOR FUNCTION MEASURE IN CEREBRAL PALSY

E. Nordmark, RPT,[1] G. Hägglund, MD, PhD,[2] and G.-B. Jarnlo, RPT, PhD[1]

*From the Departments of [1]Physical Therapy and [2]Orthopaedics, University Hospital, Lund, Sweden*

**ABSTRACT. The Gross Motor Function Measure (GMFM), an instrument comprising five dimensions devised by Russell and co-workers (7) to measure gross motor function in children with cerebral palsy (CP) or brain damage, enables changes in performance status to be evaluated after therapy or when monitored over time. We analysed its inter-rater and intra-rater reliability on the three most difficult dimensions. A video-recording of three children with CP performing test tasks was assessed on two occasions at an interval of six months by each of the 15 physiotherapists using the GMFM manual but without previous experience or training in the use of the instrument. Mean percentage scores were similar at the first and second assessments. Both inter- and intra-rater reliabilities were good, inter-rater reliability being 0.77 and 0.88 at the first and second assessments, respectively, and intra-rater reliability 0.68 at the second assessment. The findings suggest the GMFM to be a useful and reliable instrument for assessing motor function and treatment outcome in CP.**

*Key words:* cerebral palsy, children, motor function, reliability, video-recording.

## INTRODUCTION

Assessing treatment outcome in children with cerebral palsy (CP) is difficult, and diagnosis and classification of CP, as well as defining treatment in terms of the type and amount of care required are complicated. It is also difficult to distinguish between the specific effects of the specific interventions from those of overall rehabilitation efforts and the child's spontaneous improvement due to growth and natural development. Accordingly, there is a lack of satisfactory tools for assessing outcome in children with CP (1).

The Gross Motor Function Measure (GMFM) is a new standardized observational instrument designed by Russell and co-workers to yield an index of gross motor function, enabling changes in function to be evaluated after therapy or monitored over time, especially in children with CP or head injuries (7). The GMFM is based on normal gross developmental milestones, and includes tests of the following items: lying in prone and supine positions, progressing to rolling, sitting, kneeling, crawling, standing and ultimately walking, running and jumping. It has been developed for use in both clinical and research settings (6). Since its first publication in 1989, the GMFM has been spread via workshops, including training videos arranged by its constructors.

Naturally, the acid test of an instrument of this kind is its reliability, i.e. the consistency of the results obtained at repeated applications, over time or between observers (9). All measurements are susceptible to sources of variation including factors pertaining to the subject of assessment, the test, the user and the context. Russell and colleagues (8) have singled out important variables such as the performance status and degree of disability of the examinee. The duration of the assessment and the clarity of the administration guidelines are factors that may vary from one test to another. Factors associated with the environment include the test setting, temperature and time of day. Other sources of variability are patient compliance, age and background, experience of the examiner, the examiner's familiarity with the examinee, and the method of assessment, e.g. direct contact or analysis of videotaped performance (8). For reliability to be general, reliability testing needs to be conducted with people deemed to be typical users of the instrument, and subjects typical of those the instrument was designed to test. When considering the incorporation of a new test in research projects or clinical practice, it is important to determine whether the test is well standardized, valid, reliable and clinically useful (11).

It was planned to introduce the GMFM in Scandinavia, as part of the test battery for evaluating children undergoing selective dorsal rhizotomy (SDR). Accordingly, the aim of this study was to analyse the intra-observer and inter-observer reliability of the GMFM on its three most difficult dimensions at repeated administration.

## MATERIALS AND METHODS

The GMFM contains 88 items, grouped to represent five distinct areas of motor function. The five dimensions are: (A) lying and rolling, (B) sitting, (C) crawling and kneeling, (D) standing, and (E) walking, running and jumping. Each GMFM item is scored on a four-point scale, ratings of 0, 1, 2 and 3 being assigned thus: 0 = does not initiate; 1 = initiates (<10% of the task); 2 = partially completes (10 to <100% of the task); and 3 = completes the task. A percentage score can be calculated for each dimension (child's score/maximum score × 100), and an overall or aggregated score is obtained as the quotient of the sum of percentage scores divided by the number of dimensions. The responsiveness of the GMFM can be enhanced and its focus narrowed by including only dimensions where change is expected to take place, and for which target or goal scores can be set (7).

A 20-minute long video-recording was spliced together of children with CP, performing the required tasks on the 44 items from dimensions A, B and C. These three dimensions were considered the most difficult ones to score. Testing on the items was undergone by three children with CP (spastic diplegia), aged 4, 5 and 9 years, and representing different severities of disease and levels of function. This video-recording was individually observed and scored by each of 15 physical therapists working at three different rehabilitation centres in southern Sweden. The physical therapists had at least two years' experience of working with children with CP. They were all mailed introductory information on the GMFM, the introduction and the administration and scoring guidelines having been translated into Swedish from the original manual.

The guidelines contained the explicit definitions of the criteria for successful achievement of the selected items on the video-recording. The physical therapists assessed the video-recording twice at an interval of six months. None of them had used the GMFM before the first assessment or between the first and second assessments. They were also unfamiliar with the children on the video-recording.

Kendall's coefficient of concordance was used to assess inter-rater reliability, and Kendall's rank correlation to assess intra-rater reliability (10).

## RESULTS

The mean percent score was almost similar at the first assessment and at the reassessment six months later (Table I). The interrater observer reliability in the first assessment was 0.77 and at the retest 0.88. The intrarater observer reliability was 0.68 (Fig. 1).

Table I. *Mean percentage scores (mean %), ranges and SD in dimensions A, B and C of the GMFM, at the first and second assessments*

| | Dimension | Mean % | Range | SD |
|---|---|---|---|---|
| 1st assessment | A | 82 | 73–88 | 4.7 |
| | B | 84 | 77–98 | 5.9 |
| | C | 63 | 48–73 | 6.4 |
| 2nd assessment | A | 81 | 69–90 | 5.3 |
| | B | 84 | 79–94 | 4.2 |
| | C | 62 | 55–73 | 6.4 |

## DISCUSSION

The GMFM is designed to measure quantity and capacity, as distinct from quality and performance. To elicit the child's present abilities and limitations in motor function, it is important to decide whether typical performance should be measured, as opposed to the child's best-ever performance in a specific testing situation. Since both assessment and therapy are costly, constitute intrusive and disruptions of a child's life, it is important to determine why and how we assess children with CP, and what precisely we are assessing. Thus, not only do we need tools to measure and evaluate changes in motor function, but as testers it is also our responsibility to evaluate the applicability of the instruments to specific clinical situations.

The results of our study suggest the GMFM to possess both a good intra-rater and inter-rater reliability, despite the fact that only the most difficult dimensions were tested, and the observers were not trained in using the test. Earlier studies have shown the GMFM to be highly reliable, valid and sensitive to changes in a child's motor function (6). However, inter-observer reliability values for dimensions A, B and C have been reported to be lower than those for dimensions D and E (7), suggesting dimensions A, B and C to be more difficult to score.

In this study several steps were taken to maximize true response, reduce variation and measure reliability. The observers selected were experienced physical therapists, who had worked clinically with CP for at least for two years. However, as precision in results had been reported to be independent of previous clinical experience of using the GMFM (7), another selection criterion was the absence of such experience.

Physical therapists need to know whether they are using tests in a manner that produces reliable measurements, so that they can have confidence in their

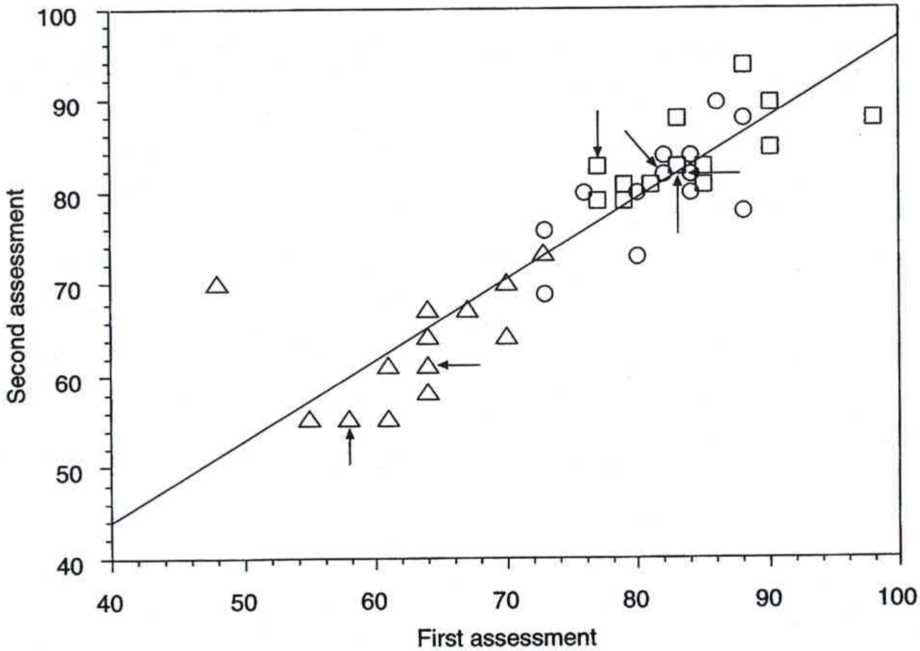


*Fig. 1.* Agreement between percentage scores by individual raters at two different assessments. ○ = Dimension A, □ = dimension B, △ = dimension C. Arrows indicate duplicated data.

ability to attribute a change in score to changes in patient function rather than to measurement error (5). Thus, for the present study a standardized test situation was required, and therefore video-recordings of children with CP performing the set tasks on dimensions A, B and C were spliced together. Using a video-recording is an efficient means of assessing patients of various ages, functional levels and severities of CP. When evaluating measures for clinical use, it is important to consider population-specific reliability for the particular group being measured, and for the type of people administrating the instrument (4). This was borne in mind when selecting children for the video-recording, all three of whom had been referred as candidates for SDR.

The video-recording was kept short (20 minutes) to avoid the impairment of the reliability due to over-long assessment. We measured test reliability on three (A, B and C) of the five dimensions, considered to be the more difficult part of the test.

There are several advantages and disadvantages of using video-recordings to capture observational data (2, 3, 8). By using a video-recording, it is possible to eliminate such confounding factors as variation in patient compliance. This advantage is particularly apparent when dealing with children. However, this approach is of no help in evaluating the physical therapist's ability to administer the instrument in practice (8). Nor does it take into account other sources of variability that are present when clinicians assess children in practice, e.g. variability due to different testers and environments. Therefore further research is needed to determine whether reliability is maintained in a clinical situation in which it is necessary to both administer and score the GMFM.

The physical therapists in our study had no previous experience of the GMFM. None the less, their ratings were characterized by a high level of agreement and a good level of agreement between the two assessments. This suggests the GMFM to be easy to use due to the helpful administration and scoring guidelines containing explicit definitions of the criteria for successful accomplishment by the examinee of the set task for each item (7). The manual was translated into Swedish in order to minimize error due to language difficulties. Inter-rater reliability was better at the reassessment, possibly due to the physical therapists being somewhat more experienced or motivated on the second occasion.

In conclusion, both the inter-rater and intra-rater reliability would seem to be good at repeated administration of the GMFM, even if the tester is not trained in its use. We suggest that the GMFM may prove to be an important measurement tool for use in follow-up to assess function and treatment outcome in CP.

## REFERENCES


1. Campbell, S. K.: Measurement of motor performance in cerebral palsy. *In* Movement disorders in children (eds H. Forssberg & H. Hirschfeld), pp. 264–271. Med Sport Sci, Karger, Basel, 1992.
2. Gross, D.: Issues related to validity of videotaped observational data. West J Nurs Res *13:* 658–663, 1991.
3. Gross, D. & Conrad, B.: Issues related to reliability of videotaped observational data. West J Nurs Res *13:* 798–803, 1991.
4. Rothstein, J. M.: Measurement and clinical practice: theory and application. *In* Measurement in physical therapy (ed. J. Rothstein), pp. 1–48. Churchill Livingstone Inc, New York, NY, 1985.
5. Rothstein, J. M., Campbell, S. K., Echternach, J. L., Jette, A. M., Knecht, H. G. & Rose, S. J.: Standards for tests and measurements in physical practice. Phys Ther *71:* 589–622, 1991.
6. Russell, D., Rosenbaum, P., Cadman, D., Gowland, C., Hardy S. & Jarvis, S.: The gross motor function measure: A means to evaluate the effects of physical therapy. Dev Med Child Neurol *31:* 341–352, 1989.
7. Russell, D., Rosenbaum, P., Gowland, C., Hardy, S., Lane, M., Plews, N., McGavin, H., Cadman, D. & Jarvis, S.: Gross motor function measure manual. 2nd ed. Hamilton, McMaster University, Canada, 1993.
8. Russell, D., Rosenbaum, P., Lane, M., Gowland, C., Goldsmith, C., Boyce, W. & Plews, N.: Training users in gross motor function measure: methodological and practical issues. Phys Ther *74:* 630–636, 1994.
9. Sackett, D. L., Haynes, R. B., Gyatt, G. H. & Tugwell, P.: Clinical epidemiology: A basic science for the clinical medicine. 2nd ed. Little Brown and Co. Inc., Boston, Mass, 1991.
10. Siegel, S.: Non-parametric statistics for the behavioral sciences. McGraw Hill Book Company, New York, 1956.
11. Stengel, T. J.: Assessing motor development in children. *In* Pediatric neurologic physical therapy (ed. S. K. Campbell), pp. 33–65. New York, Churchill Livingston Inc., 1991.


*Accepted May 25, 1996*

*Address for offprints:*

Eva Nordmark, RPT
Department of Physical Therapy
University of Lund
PO Box 5134
SE-220 05 Lund, Sweden