J Rehabil Med 2008; 40: 589–591

SHORT COMMUNICATION

HIGH INTER-TESTER RELIABILITY OF THE NEW MOBILITY SCORE IN PATIENTS WITH HIP FRACTURE

Morten Tange Kristensen, PT[1,2,3], Thomas Bandholm, MSc, PT[4], Nicolai Bang Foss, MD[5], Charlotte Ekdahl, RPT, PhD[1] and Henrik Kehlet, MD, PhD[6]

From the [1]Department of Health Sciences, Division of Physiotherapy, Lund University Hospital, Lund University, Lund, Sweden, [2]Department of Physiotherapy, [3]Orthopaedic Surgery, [4]Gait Analysis Laboratory, Department of Orthopaedic Surgery, [5]Department of Anaesthesiology, Hvidovre University Hospital and [6]Section of Surgical Pathophysiology, Rigshospitalet, Copenhagen University, Copenhagen, Denmark

Objective: **To assess the inter-tester reliability of the New Mobility Score in patients with acute hip fracture.**
Design: **An inter-tester reliability study.**
Subjects: **Forty-eight consecutive patients with acute hip fracture at a median age of 84 (interquartile range, 76–89) years; 40 admitted from their own home and 8 from nursing homes to an acute orthopaedic hip fracture unit at a university hospital.**
Methods: **The New Mobility Score, which evaluates the prefracture functional level with a score from 0 (not able to walk at all) to 9 (fully independent), was assessed by 2 independent physiotherapists at the orthopaedic ward. Inter-tester reliability was evaluated using the intraclass correlation coefficient ($ICC_{1,1}$) and the standard error of measurement (SEM).**
Results: **The ICC between the 2 physiotherapists was 0.98, 95% confidence interval (CI) 0.96–0.99 and the SEM was 0.42, 95% CI –0.40–1.24 New Mobility Score points. No systematic between-rater bias was observed ($p > 0.05$). Patients who were scored differently by the 2 physiotherapists had significantly lower mental scores ($p = 0.02$).**
Conclusion: **The inter-tester reliability of the New Mobility Score is very high and can be recommended to evaluate the prefracture functional level in patients with acute hip fracture.**

Key words: hip fracture, activities of daily living, rehabilitation, reproducibility of results.

J Rehabil Med 2008; 40: 589–591

Correspondence address: Morten Tange Kristensen, Department of Physiotherapy 236 and Orthopaedic Surgery, Hvidovre University Hospital, Kettegaard Alle 30, Copenhagen DK-2650, Denmark. E-mail: morten.tange.kristensen@hvh.regionh.dk

Submitted January 3, 2008; accepted March 26, 2008

INTRODUCTION

The New Mobility Score (NMS) (1) is a validated predictor of long-term mortality and rehabilitation outcome in patients with hip fracture (1, 2). The score has been used to stratify patients with acute hip fracture according to functional capacity (3) and to describe the pre-fracture functional level (4). The NMS is a composite score of a patient's ability to perform: indoor walking, outdoor walking and shopping pre-fracture. It provides a score between 0 and 3 (0: not at all, 1: with help from another person, 2: with an aid, 3: no difficulty) for each function, resulting in a total score from 0 (no walking ability at all) to 9 (fully independent). Parker & Palmer (1) found a cut-off point at 5 to be the best predictor of 1-year mortality after hip fracture. Other studies (2, 5) have used or examined the NMS with a single cut-off point at 5, dividing the patient population into 2 groups (NMS 0–5 vs 6–9). As different physiotherapists or physicians usually obtain the score, it is important to establish the inter-tester reliability. No such data have been reported. Reliability refers to the consistency of a test or measurement (6) and it can be quantified as either relative or absolute (6, 7). Relative reliability is often expressed by the intraclass correlation coefficient (ICC), which indicates the relationship between 2 or more measures of the same test or score, with a coefficient from zero to one (6, 7). The standard error of measurement (SEM), which quantifies the precision of individual scores on a test, gives a clinician a result in the same unit as the measurement (6, 8), thereby indicating whether a change in score is a real change (7). The purpose of this study was to determine the relative and absolute inter-tester reliability of the NMS in patients with acute hip fracture.

PATIENTS AND METHODS

Participants were 48 consecutive patients (40 from their own home; median age (25–75 quartiles), 81 (75–86) years and 8 from nursing home; age 91 (88–93) years) admitted to a specialized 14-bed hip-fracture unit. This study is part of Hvidovre University Hospitals hip fracture project that has been approved by the local ethics committee and by the Danish data protection agency.

Information on age, gender and a validated 9-point Danish version of the abbreviated mental test score was taken upon admission (9). The assessment of the NMS relies on the individual's ability to recall their prefracture functional level. Therefore, to avoid recall bias the NMS was obtained by 2 independent physiotherapists (PTA and PTB) at different days post-surgery.

Statistics

Descriptive statistics and correlations (Spearman's rho) for all patients were calculated for age, gender, prefracture functional level (using

NMS) and mental status on admission. Systematic between-rater bias was assessed using the Mann-Whitney *U* test. Relative reliability was calculated using the $ICC_{1.1}$ with the corresponding 95% confidence interval (95% CI). Absolute reliability was calculated as the SEM using the equation standard deviation (SD)$\sqrt{1}$–ICC, where SD is the SD of the NMS-scores from all patients (6). The 95% CI was calculated for the SEM as $\pm 1.96 \times$ SEM. Scatters of the between-rater differences were plotted against the rater means (10) to indicate if the between-rater differences were related to the NMS-score (heteroscedasticity). This was not the case as no significant relationship between the numerical between-rater differences and rater means was observed ($r = -0.051$, $p = 0.733$, Spearman's rho). Finally, the number of patients with between-rater differences in total NMS-scores was calculated for all score differences. All analyses were conducted using SPSS for windows version 11.5. The level of significance was set at $p < 0.05$.

RESULTS

The ICC was 0.98 (95% CI 0.96–0.99), while the SEM was 0.42 NMS points (95% CI –0.40–1.24, Fig. 1) and no systematic between-rater bias was observed ($p > 0.05$). The NMS score (mean of PTA and PTB) was significantly ($p < 0.001$) correlated with age ($r = -0.584$) and mental scores on admission ($r = 0.612$), and women were significantly older than men ($p = 0.014$)**.** The NMS was obtained at median (25–75 quartiles) 1.5 day (1–2) and 3 days (2–6) post-surgery. Scores between PTA and PTB differed in 7 out of 48 patients (14%) (Table I). These 7 patients had significantly lower mental scores ($p = 0.02$). Only 2 out of these 7 scores differed more than 1 point, and NMS score differences were not related to the interval in days between the first and second NMS assessment ($p = 0.682$, Mann-Whitney *U* test).

DISCUSSION

The present study showed a high inter-tester reliability of the NMS, with only 2 out of 48 (4%) recorded scores differing more than one point. A significant correlation between age, mental scores, and NMS scores were found for all patients, and patients with NMS score-differences had lower mental scores than patients with equally recorded NMS-scores. There is no clear definition on the interpretation of the ICC, but Munro et al. (11) describe correlations from 0.90 and above as being "very high". Therefore, the reliability of the NMS in this study, when evaluated by the ICC, was very high (0.98) and the true NMS score for a patient with a recorded score of 4 will, with a 95% CI of ± 0.82, only diverge by 1 point. We chose not to investigate the intra-tester reliability of the NMS in the present study, as each physiotherapist was likely to remember the answers of the first NMS recording (recall bias). In addition, it is a common finding that the intra-tester reliability is higher compared with the inter-tester reliability (12, 13).

Previous studies (1, 2) have found the NMS to be a valid predictor of mortality and rehabilitation outcome and findings from the present study suggest that the relative and absolute inter-tester reliability of the NMS is very high. That is, different personnel can record the NMS with a high possibility of obtaining the same score. Ward personnel should be careful when recording the score in patients with lower mental scores.

In conclusion, we recommend the NMS to evaluate the pre-fracture functional level in patients with acute hip fracture.

Table I. *Absolute differences in total scores of the New Mobility Score (NMS) in 48 hip fracture patients*

Differences in NMS points	Subjects		Cumulative percentage
	n	%	
PTA vs PTB			
0	41	86	86
1	5	10	96
2	1	2	98
3	1	2	100

PTA: physiotherapist A; PTB: physiotherapist B.

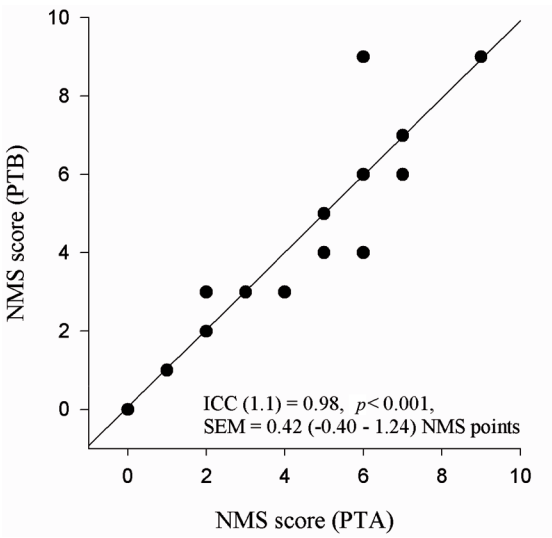

Fig. 1. Relationship between the New Mobility Score (NMS) obtained by 2 physiotherapists (PTA) and (PTB) in 48 patients with hip fracture. ICC: intraclass correlation coefficient; SEM: standard error of measurement.

ACKNOWLEDGEMENTS

This project was funded by a grant from the IMK Fonden, Copenhagen, Denmark.

REFERENCES

1. Parker MJ, Palmer CR. A new mobility score for predicting mortality after hip fracture. J Bone Joint Surg Br 1993; 75: 797–798.
2. Kristensen MT, Foss NB, Kehlet H. Timed up & go og New Mobility Score til prædiktion af funktion seks måneder efter hoftefraktur [Timed up and go and new mobility score as predictors of function six months after hip fracture]. Ugeskr Laeger 2005; 167: 3297–3300 (in Danish).
3. Foss NB, Kristensen MT, Kristensen BB, Jensen PS, Kehlet H. Effect of postoperative epidural analgesia on rehabilitation and pain after hip fracture surgery: a randomized, double-blind, placebo-controlled trial. Anesthesiology 2005; 102: 1197–1204.
4. Kristensen MT, Foss NB, Kehlet H. Timed "up & go" test as a predictor of falls within 6 months after hip fracture surgery. Phys Ther 2007; 87: 24–30.

5. Foss NB, Kristensen MT, Kehlet H. Prediction of postoperative morbidity, mortality and rehabilitation in hip fracture patients: the cumulated ambulation score. Clin Rehabil 2006; 20: 701–708.
6. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005; 19: 231–240.
7. Sole G, Hamren J, Milosavljevic S, Nicholson H, Sullivan SJ. Test-retest reliability of isokinetic knee extension and flexion. Arch Phys Med Rehabil 2007; 88: 626–631.
8. Flansbjer UB, Holmback AM, Downham D, Patten C, Lexell J. Reliability of gait performance tests in men and women with hemiparesis after stroke. J Rehabil Med 2005; 37: 75–82.
9. Qureshi KN, Hodkinson HM. Evaluation of a ten-question mental test in the institutionalized elderly. Age Ageing 1974; 3: 152–157.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1: 307–310.
11. Munro BH, Visintainer MA, Page EB, editors. Statistical methods for health care research. Philadelphia: JB Lippincott; 1986.
12. Pervez H, Parker MJ, Pryor GA, Lutchman L, Chirodian N. Classification of trochanteric fracture of the proximal femur: a study of the reliability of current systems. Injury 2002; 33: 713–715.
13. Karagiannopoulos C, Sitler M, Michlovitz S. Reliability of 2 functional goniometric methods for measuring forearm pronation and supination active range of motion. J Orthop Sports Phys Ther 2003; 33: 523–531.