

SHORT COMMUNICATION

COULD MOBILITY PERFORMANCE MEASURES BE USED TO EVALUATE WHEELCHAIR SKILLS?

Didier Pradon, PhD¹, Nicolas Pinsault, PT, PhD², Raphaël Zory, PhD¹ and François Routhier, PhD³

From the ¹Groupement de Recherche Clinique et Technologique sur le Handicap (EA 4497), CIC-IT 805, CHU Raymond Poincaré, Garches, ²Ecole de kinésithérapie du Centre Hospitalier Universitaire, Grenoble, France and ³Centre interdisciplinaire de recherche en réadaptation et intégration sociale (CIRRS), Institut de réadaptation en déficience physique de Québec (IRDPQ), Québec, Canada

Objective: To determine the relationship between mobility performance measures and Wheelchair Skill Test (WST) scores and to establish the test-retest and inter-rater reliability of these measures.

Methods: Forty patients with spinal cord injury participated in this study. Subjects performed the Wheelchair Skill Test and mobility performance tests: maximal velocity (V_{max}), spontaneous velocity (V_{spont}) and a 10-m back and forth slalom (S_{time}). Eighteen patients with spinal cord injury participated in a second testing session to evaluate test-retest reliability and, among these patients, 8 participated in a third testing session to evaluate inter-rater reliability.

Results: Spearman's correlation coefficients calculated between WST and V_{max} , V_{spont} and S_{time} were high and significant ($p < 0.05$). The intraclass correlation coefficients (ICC (2.1)) evaluating test-retest reliability for V_{max} , V_{spont} and S_{time} were 0.94, 0.84 and 0.88, respectively. The ICC evaluating inter-rater reliability for V_{max} , V_{spont} and S_{time} were 0.92, 0.92 and 0.95, respectively. Reliability results were confirmed by Bland-Altman plots.

Conclusion: V_{max} and S_{time} could be used to evaluate wheelchair skills and to create a new scale, whereas V_{spont} is the least appropriate of these measurements to describe wheelchair skills.

Key words: spinal cord injury; evaluation; Wheelchair Skills Test; Spearman's correlation.

J Rehabil Med 2012; 44: 276–279

Correspondence address: Didier Pradon, EA 4497 GRTCH, CIC-IT 805, Laboratoire d'analyse de la marche – Hôpital Raymond Poincaré, 104 Bd Raymond Poincaré, FR-92380 Garches, France. E-mail: didier.pradon@rpc.aphp.fr

Submitted May 2, 2011; accepted October 21, 2011

INTRODUCTION

The Council of Europe estimated in 2002 that by 2005 more than 2.5 million people worldwide could be living with spinal cord injury (SCI) (1). Whereas it is difficult to establish an accurate picture of the epidemiology of spinal cord injuries (SCI), more recent studies are in line with the Council of Europe estimate, and show that the prevalence of SCI is increasing (2, 3). Most patients with SCI need rehabilitation and physiotherapy to help

them maximize their potential and learn to live efficiently as wheelchair users. To enable this process, the objective evaluation of manual wheelchair skills is of great importance.

There are some excellent instruments to evaluate manual wheelchair skills described in the literature (4–6). The Wheelchair Skill Test (WST), which was initially proposed by Kirby et al. (6) in 2002 and subsequently revised systematically (7), has the advantage of evaluating manual wheelchair function at the level of activities of daily living, and demonstrates good metrological properties (8). In fact, establishing the reliability and validity of a given evaluation constitutes an important, but not sufficient, prerequisite for its use and for the interpretation of collected parameters. For example, practical considerations, such as applicability (9), which is the quality of a tool that enables its use with a given population or in a specific context, appear to be an important complement to metrological properties. In fact, the WST is traditionally mentioned as the standard for these objective evaluations, but it sometimes fails to reach the standards required for widespread use (10). Different wheelchair or mobility performance tests have been developed. For example, a 9-task wheelchair circuit (including a sprint, figure-of-8 shape, transfer, etc.) has been proposed and validated by Kilkens et al. (11). This kind of test presents very good reliability. However, the applicability of the test under clinical conditions is more difficult because it requires time and equipment. At the same time, many simple outcomes are collected during adapted physical activity or physiotherapy sessions that could expand on fill out the WST evaluation and make the wheelchair selection process easier for clinicians and their patients. Among these outcomes, we can take into account the mobility performance measures, corresponding to the wheelchair users' ability to move easily and efficiently, which bring together wheeling performance (expressed through speeds) and handiness (expressed through the time taken to perform a slalom).

The aim of the present study was to determine the relationship between mobility performance measures and WST scores and to establish the test-retest and inter-rater reliability of these measures.

METHODS

Subjects

Forty SCI patients participated in the first testing session (30 males and 10 females; mean age 36.9 years (standard deviation (SD) 11.2); mean

height 172.8 cm (SD 9.9); mean weight 68.6 kg (SD 12.5)). Within these subjects, 18 could be classified as low paraplegics (level of SCI from T10–L3), 15 as high paraplegics (T1–T9) and 7 as tetraplegics (C6–C7). We used a heterogeneous sample in order that the sample was as close as possible to clinical realities and interrogations. All the SCI patients were, or had been, involved in a rehabilitation programme. The subjects had a mean of 79.8 months' experience (range 1–360). There was no attempt to stratify the sample on the basis of age, sex or level of SCI. The study inclusion criteria were: 20 years of age or older; cooperative and pain-free; competent to give informed consent; and willing to participate. Prior to involvement in this study, each participant signed a letter indicating informed consent. The study was conducted according to the Declaration of Helsinki and approval for the project was obtained from the institutional ethics committee.

Experimental procedure

The 40 patients with SCI participated in the first testing session (S1). First, subjects performed the WST (3.2 version). The WST was administered as outlined in the WST test manual (7). WST provides capacity scores for each skill based on explicit criteria. Each participant was evaluated in the wheelchair that they were using on the day of the study and equipped in his or her usual manner when using the wheelchair. The total scores were calculated according to the WST 3.2 manual (7).

Next, subjects performed the 3 wheelchair mobility tests: (i) maximal velocity (V_{\max} ; km/h); (ii) spontaneous velocity (V_{spont} ; km/h); and (iii) slalom (S_{time} ; s). For V_{\max} measurement, subjects were instructed to wheel the chair in a 20-m straight line at the fastest speed they could reach. Value of V_{\max} was obtained for the first 12 m only. Three trials were completed, and the fastest was accepted as the measure of V_{\max} . V_{spont} was obtained by asking subjects to wheel the chair in a 20-m straight line at their preferred speed. The value of V_{spont} was obtained for the first 12 m only. Three trials were completed and their mean was accepted as the measure of V_{spont} . To evaluate the handiness of the wheelchair, subjects performed a 10-m back and forth slalom (each block separated by 1 m) as fast as possible. The measure (slalom) was the total time (S_{time} ; s) needed to complete the task considering that each missed (or felt) block added 10 s to the total slalom time.

Eighteen SCI patients agreed to participate in a second experiment (S2, 15 males and 3 females) and, among these patients, 8 participated in a third experiment (S3, 5 males and 3 females). S2 and S3 were similar to S1. For S2, the entire procedure evaluating wheelchair mobility performance was repeated by the same rater as for S1, on the 18 patients after a 1-week interval to evaluate test-retest reliability. For S3, two different raters administered the entire procedure, evaluating wheelchair mobility performance for 8 patients within the same day to evaluate inter-rater reliability. Values obtained for each experimenter were not communicated to the other.

Statistical analysis

All statistical calculations were carried out with R 2.0.1 software (R Core Development Team). Means, SD, range and standard errors were calculated for each parameter. A level of $p < 0.05$ was used to identify statistical significance. Considering that the data do not appear to follow a normal distribution, the relationship between WST scores and V_{\max} , V_{spont} and S_{time} was established using a Spearman's rank correlation coefficient (Rs). The Rs values were interpreted according to Domholdt recommendations (12). Spearman's rank correlation coefficients were calculated for: (i) total sample ($n = 40$); (ii) the 20 subjects with the worst WST score (G1, $n = 20$); and (iii) the 20 subjects with the best WST score excluding patient with the maximal score (G2, $n = 20$).

The test-retest and inter-rater reliability of V_{\max} , V_{spont} and S_{time} were assessed with the intraclass correlation coefficient (ICC (2,1)) (13) and according to Fleiss' classification (14). The standard error of measurement (SEM) and the 95% confidence interval (CI) of ICC values were calculated for all dependent variables (15). The use of the 95% CI demonstrates how closely the measurements agree on different

occasions, whereas the SEM indicates the precision of measurements. Finally, Bland-Altman graphs were formed to give a visual interpretation of the data as well as to determine potential bias (16).

RESULTS

The total group of subjects obtained a mean WST score of 83.8% (SD 16.6%) (range 47.0–100.0%), a mean V_{\max} of 6.91 km/h (SD 2.10) (range 3.6–11.7), a mean V_{spont} of 4.74 km/h (SD 1.21) (range 2.6–7.5) and a mean S_{time} of 60.4 s (SD 29.6) (range 25.0–153.0). For the total sample and G1, Spearman's rank correlation coefficients between WST score and mobility performance measurement (V_{\max} , V_{spont} and S_{time}) were all significant (Table I). According to Domholdt recommendations (12), the correlation between WST score and V_{\max} and S_{time} is high, whereas the correlation between WST and V_{spont} is moderate. Spearman's rank correlation coefficients calculated on G2 ($n = 20$) are all very low and non-significant.

Concerning the test-retest reliability ($n = 18$), analysis of V_{\max} showed an ICC of 0.94 (95% CI 0.86–0.98, SEM 0.15). For V_{spont} , the ICC value was 0.84 (95% CI 0.68–0.91, SEM 0.26) and for S_{time} the ICC value was 0.88 (95% CI 0.72–0.98, SEM 2.92). Finally, for the inter-rater reliability ($n = 8$), the ICC value was 0.92 for V_{\max} (95% CI 0.60–0.98, SEM 0.17), 0.92 for V_{spont} (95% CI 0.15–0.99, SEM 0.11), and 0.95 for S_{time} (95% CI 0.82–0.98, SEM 1.35). In addition to those good to excellent ICCs, Bland-Altman plots showed no specific or major trends between testers or between testing time (Figs 1 and 2).

DISCUSSION

The aim of the present experiment was to determine the relationship between wheelchair mobility tests and WST scores and the test-retest reliability and inter-rater reliability of these measures. The key findings of this study were that: (i) the relationship between mobility performance test and WST score are high when considering patients with the lowest score at the WST (G1) or the total sample; and (ii) the test-retest reliability of the 3 mobility performance tests (V_{\max} , V_{spont} and S_{time}) is excellent.

The WST scores of the 40 SCI patients were first compared with wheelchair mobility performance measures such as V_{\max} , V_{spont} and S_{time} . Firstly, our results showed that V_{\max} and S_{time} were "highly" correlated with WST scores, whereas V_{spont} was only "moderately" correlated with WST scores. In other words, the

Table I. Spearman's rank correlation coefficient between Wheelchair Skill Test score and the maximal velocity (V_{\max}), the spontaneous velocity (V_{spont}) and the slalom time (S_{time}) for the total sample ($n = 40$), group 1 ($n = 20$) and group 2 ($n = 20$)

Variables	Total	Group 1	Group 2
V_{\max}	0.72*	0.80*	0.31
V_{spont}	0.57*	0.64*	0.17
S_{time}	-0.75*	-0.75*	-0.33

*Spearman's rank correlation coefficient were significant at $p < 0.05$.

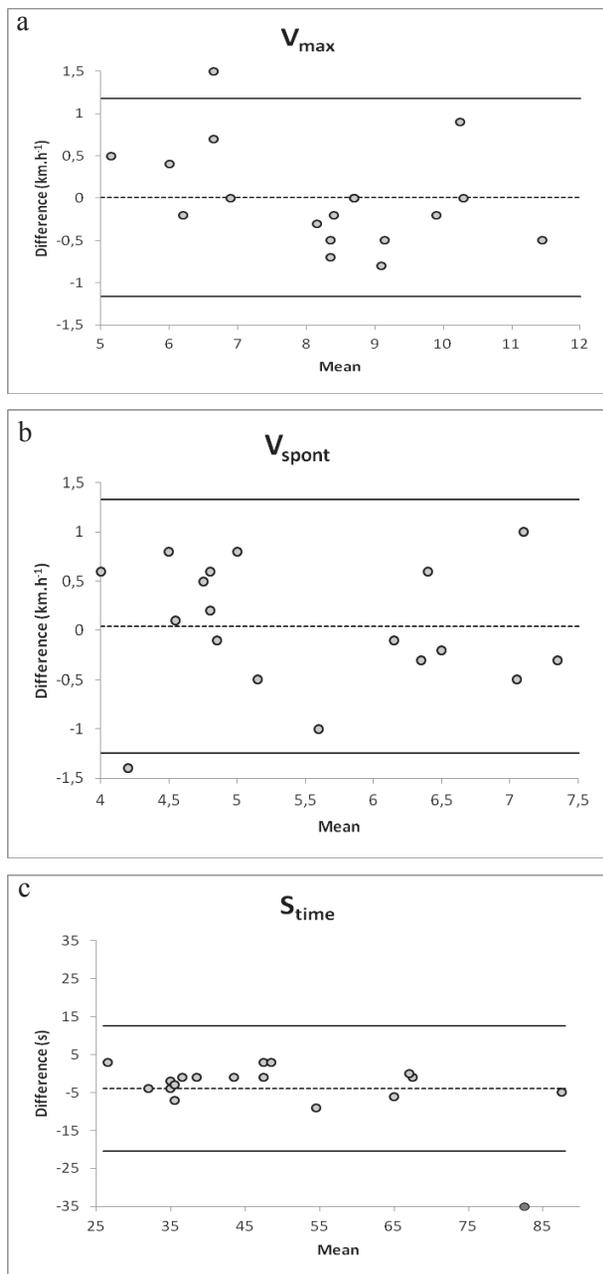


Fig. 1. Test-retest Bland-Altman plots: difference against mean for (a) maximal velocity V_{max} ; (b) spontaneous velocity V_{spont} ; (c) slalom time S_{time} . Dotted line: mean. Solid line: limits of agreement according to mean \pm 2 standard deviations.

results suggest that V_{spont} is the least appropriate of the measures in this study to describe wheelchair skills. This result is not surprising considering that the instruction “to wheel at preferred speed” is highly subjective for the subjects and very difficult for the clinician to interpret. The high level of correlation found between WST and V_{max} and S_{time} also underlines the good general validity of the WST. However, the very low Spearman’s rank correlation coefficient observed between WST scores and the mobility performance tests for G2 (the best scores on WST, $n=20$) highlighted

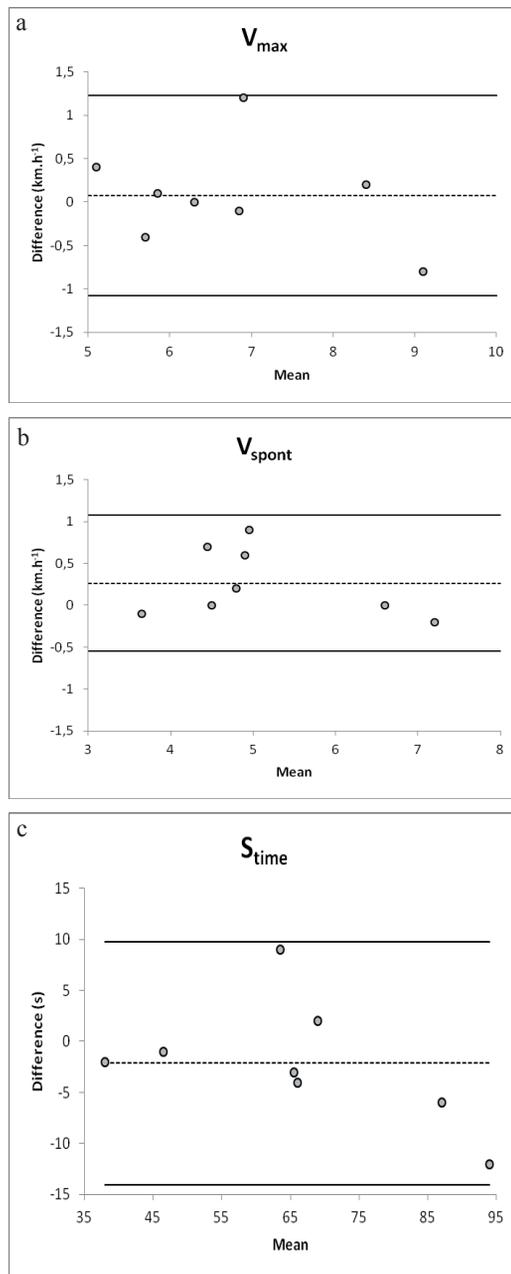


Fig. 2. Inter-rater Bland-Altman plots: difference against mean for (a) maximal velocity V_{max} ; (b) spontaneous velocity V_{spont} ; (c) slalom time S_{time} . Dotted line: mean. Solid line: limits of agreement according to mean \pm 2 standard deviations.

an important limit of the WST. Indeed, this result shows that the WST has difficulty in differentiating patients with a high level of performance. Although the WST has the advantage of being applicable to all kinds of wheelchair users, when we consider the application of this test for specific users, such as patients with SCI, there are some difficulties in discriminating patients. In fact, sportsmen with SCI generally obtained maximum or very high scores on the WST, whereas clinically they might exhibit very different skill and mobility performances.

Secondly, our results showed that all parameters demonstrated excellent test-retest and inter-rater reliability. For V_{\max} , these results confirm those of a recent study that highlighted high test-retest and inter-rater reliability on a similar test (13). Nevertheless, the reliability is necessary, but not sufficient, for an evaluation to be used. In fact, Mortensen et al. (10) underlined the importance of the applicability. On this particular point, the evaluation of wheelchair mobility performance measures takes less than 5 min. These measures could be a good counterpart to the WST, which is much more time-consuming and perhaps less discriminative. Taken together, these results suggest that V_{\max} and S_{time} could be used in clinical practice and integrated in a new scale that is reliable, easy to use and not time-consuming.

It is important to note that this study focused only on SCI patients, whereas many other persons use manual wheelchairs. However, focusing on this group enabled us to evaluate a wide range of wheelchair skills. Moreover, because of personal constraints, only a few patients participated in experiments evaluating the reliability of the measures, thus widening the 95% CI of the ICC and limiting the impact of our results. Future research will study more subjects in order to develop a new scale that is easy to use, is not time-consuming and is based on performance measures.

ACKNOWLEDGEMENT

The authors would like to thank the all of the subjects for their time and for enduring the rigors of the protocol and data acquisition process.

REFERENCES

1. Parliamentary Assembly. Towards concerted efforts for treating and curing spinal cord injury. Doc. 9401: Report. Social, Health and Family Affairs Committee. 2002 Mar 27.
2. Hagen EM, Eide GE, Rekand T, Gilhus NE, Gronning M. A 50-year follow-up of the incidence of traumatic spinal cord injuries in Western Norway. *Spinal Cord* 2010; 48: 313–318.
3. Ahoniemi E, Alaranta H, Hokkinen EM, Valtonen K, Kautiainen H. Incidence of traumatic spinal cord injuries in Finland over a 30-year period. *Spinal Cord* 2008; 46: 781–784.
4. Routhier F, Vincent C, Desrosiers J, Nadeau S. Mobility of wheelchair users: a proposed performance assessment framework. *Disabil Rehabil* 2003; 25: 19–34.
5. Kilkins OJ, Post MW, van der Woude LH, Dallmeijer AJ, van den Heuvel WJ. The wheelchair circuit: reliability of a test to assess mobility in persons with spinal cord injuries. *Arch Phys Med Rehabil* 2002; 83: 1783–1788.
6. Kirby RL, Swuste J, Dupuis DJ, MacLeod DA, Monroe R. Wheelchair Skills Test: pilot study of a new outcome measure. *Arch Phys Med Rehabil* 2002; 83: 10–18.
7. Dalhousie University Faculty of Medicine. Wheelchair Skills Program [Cited 2010 Sept 20]. Available from: <http://www.wheelchairskillsprogram.ca/>.
8. Kirby RL, Dupuis DJ, Macphee AH, Coolen AL, Smith C, Best KL, et al. The Wheelchair Skills Test (version 2.4): measurement properties. *Arch Phys Med Rehabil* 2004; 85: 794–804.
9. Wade DT. Assessment, measurement and data collection tools. *Clin Rehabil* 2004; 18: 233–237.
10. Mortenson WB, Miller WC, Auger C. Issues for the selection of wheelchair-specific activity and participation outcome measures: a review. *Arch Phys Med Rehabil* 2008; 89: 1177–1186.
11. Kilkins OJ, Post MW, van der Woude LH, Dallmeijer AJ, van den Heuvel WJ. The wheelchair circuit: reliability of a test to assess mobility in persons with spinal cord injuries. *Arch Phys Med Rehabil*. 2002; 83: 1783–1788.
12. Domholdt E. Physical therapy research: principles and application, 2nd ed. Philadelphia: WB Saunders Co.; 2000.
13. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; 1: 30–46.
14. Fleiss JL. The design and analysis of clinical experiments. New York: Wiley; 1986.
15. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther* 1997; 77: 745–750.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327: 283–338.