ELSEVIER

Parkinsonism and Related Disorders 11 (2005) 19–24

Parkinsonism & Related Disorders

www.elsevier.com/locate/parkreldis

Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study

L.I.I.K. Lim[a,*], E.E.H. van Wegen[a], C.J.T de Goede[a], D. Jones[b], L. Rochester[b], V. Hetherington[b], A. Nieuwboer[c], A.M. Willems[c], G. Kwakkel[a]

[a]*Department of physiotherapy, VU University Medical Center (VUmc), Amsterdam, The Netherlands*
[b]*School of Health, Community and Education Studies, Northumbria University, Newcastle upon Tyne, UK*
[c]*Department of Rehabilitation Sciences, Faculty of Physical Education and Physiotherapy, Katholieke Universiteit Leuven, Belgium*

Received 12 February 2004; revised 10 June 2004; accepted 18 June 2004

Abstract

The aim of this study was to assess reliability, responsiveness and feasibility of gait and gait related tests in the home of patients with Parkinson's disease (PD). The Unified Parkinson's Disease Rating Scale, a timed walking test, the Timed Get Up and Go test the Berg Balance Scale and the Functional Reach test were applied by three independent observers on 26 PD patients. Moderate to high Intraclass Correlation Coefficients were found, ranging from 0.74 to 0.88 and 0.64 to 0.87 for the intra- and inter-observer reliability, respectively. All test showed Reliable Change Indexes under 11% and the whole test battery was applicable within 30 min.

Keywords: Parkinson disease; Gait; Reproducibility of results; Outcome assessment

1. Introduction

Parkinson's disease (PD) is a chronic, progressive disorder characterized by movement-related symptoms such as bradykinesia, tremor, rigidity, freezing and festination. Consequently patients experience problems in gait and in gait- related activities such as balance control and transfers [1]. The consensus view is that if one wants to build an overall picture of the gait-related problems that patients with PD experience on a daily basis, a range of measurements should be used [2–5] and therefore, a comprehensive battery of tests is needed. Such a test battery should include measures that reflect the 'domain of activities' according to the International Classification of Functioning, Disability and Health (ICF) [6]. The selection of tests should be based on methodological considerations regarding reliability and validity.

Performance tests to assess gait-related functioning in PD recommended in the literature include: the Unified Parkinson's Disease Rating Scale—motor examination section (UPDRS-III) [7,8], the Timed Get Up and Go test (TGUG) [9], the Berg Balance Scale (BBS) [10], the Functional Reach test (Fr) [10] and timed walking tests [11,12]. The full UPDRS is recommended to assess disease severity [2,3,13–15]. It has been suggested that patients with PD should preferably be treated and tested in their own home situation [16,17]. However, the reliability of these tests aimed at assessing gait and balance control has never been established in the patients' own home situation, where the circumstances are less optimal for standardization compared to a clinical setting. Therefore, the aim of the present study was to assess the reproducibility, responsiveness and feasibility of performance tests used to measure gait and gait-related problems in the patients' own home environment. A gait-related test battery was compiled for patients with PD in collaboration with three different countries within Europe. This test battery will be used in a large randomised multicenter trial (RCT) aimed at investigating the effects of cueing strategies on gait and gait-related activities in the patients' own home situation [18].

* Corresponding authorTel.: +31-20-4440463; fax: +31-20-4440469.
E-mail address: i.lim@vumc.nl (L.I.I.K. Lim).

1353-8020/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.parkreldis.2004.06.003

2. Methods

2.1. Patient selection

Twenty-six subjects (15 male, 11 female) with a diagnosis of idiopathic PD were included in the study. The average age was 62.5 years (range 44–80 years) and average duration of symptoms was 6.5 years (range 1–20 years). Inclusion criteria were: (1) a Hoehn and Yahr [19] stage ranging from 1 to 3; (2) a stable medication regime; (3) ability to walk independently without a walking device; (4) absence of co-morbidity that may influence mobility; (5) an age under 80 years; (5) sufficient orientation in time and place (Mini Mental State Examination, MMSE, $=24$ [20]) and (6) completion of an informed consent for participation.

All PD-patients were recruited from the VU University Medical Center (VUmc) and lived in a geographically defined district around Amsterdam. The study was approved by the ethics committee of the VUmc.

2.2. Measurements

The test battery was compiled to assess severity of PD and gait-related functioning. The test battery consisted of the UPDRS, including the motor examination section (UPDRS III) [2], the TGUG test [21], the timed 10 m walk test (10mw), the BBS [22] and the Fr test [23]. In order to obtain uniformity of assessment practical guidelines were developed which included instructions on how to deal with specific difficulties inherent to testing in patients' homes, such as lack of space and obstacles that could impede proper gait and balance assessment.

2.3. Observers

Three physiotherapists were employed to perform the assessments. Prior to testing, the observers were trained to apply the test battery in a uniform way in three healthy subjects and one patient with PD. In addition, they were instructed on how to use the practical guidelines. The training of the three observers was conducted by two coaches (i.e. a physiotherapist (CG) and a human movement scientist (EW)) who were experienced in clinical research in PD.

2.4. Design

Patients were assessed by the three independent observers during two consecutive visits to the patients' home. All subjects were visited at approximately the same time of the day (maximum difference in time was about 1 h) to minimize existing circadian fluctuations. All patients were measured in the 'on-phase' about 1 h after medication intake. Before assessment, randomisation was applied for the sequence of the tests as well as for the sequence in which the three observers assessed the patient.

2.5. Data analysis and statistics

The inter-observer reliability for the three observers was calculated using Intraclass Correlation Coefficients (ICC) using a two-way random effects model with an absolute agreement definition [24]. Similarly, the intra-observer reliability for one of the three assessors was determined by applying a two-way mixed effects model for absolute agreement [24] and by using the Bland and Altman method [25]. ICC's were preferred because this statistic is able to deal with dichotomous outcomes, corrects for systematic errors and can be used for more than two observers [24]. The use of ICC's and Bland and Altman method gives complementary information as shown by Rankin and Stokes [26]. For the Bland and Altman method the 'limits of agreement' were computed, defined as $\pm 1.96 \times$ standard deviation of the difference score. Assuming a normal distribution of the found differences, only 95% of the differences between two measurements per individual in a stable population will be between the limits of agreement [25].

The responsiveness of the tests was determined using the Smallest Detectable Difference (SDD), the SDD was calculated on the basis of the standard error of measurement [27] (see Eq. (1)), assuming that the measurement errors were constant across the range of possible scores [28].

$$\mathrm{SDD} = \mathrm{SEM} \times 1.96 \times \sqrt{2} \quad (1)$$

where SDD is smallest detectable difference, and SEM, standard error of measurement.

In order to allow comparison of responsiveness between the tests, the Reliable Change Index (RCI) was determined for each measurement by calculating the SDD as the percentage of the maximal feasible score. Each hypothesis was tested with a two-tailed analysis with 0.05 as the level of significance.

Feasibility was determined by measuring the time needed to apply the whole test battery, including the time needed to adapt the home environment for assessments (e.g. moving furniture).

3. Results

The patients in this study had a H-Y- score ranging from 1 to 3 patient characteristics are shown in Table 1. The duration of the first visit was approximately 45 min, whereas the duration of the second visit ranged from 30 to 35 min. The median number of days between two test sessions of the same observer was 7 (inter quartile range 3 days).

3.1. Inter- and intra-observer reliability

Table 2 shows the reliability between and within observers. ICC's for *inter*- observer reliability ranged from 0.64 for the Fr test to 0.87 for the mean walking speed in the 10 m walk test. ICC's for *intra*- observer

Table 1
Patient characteristics

Variable	Mean (SD)	Range
Age (years)[a]	62.5 (8.2)	44–80
Disease Duration (years)[a]	6.5 (4.2)	1–20
Gender (M/F)[b]	15/11	
Medication[b]		
None	4	
Levodopa (in combination with benzeraside or carbidopa)	14	
Dopamine agonists	18	
Selegiline	4	
Parasympathicoliticum	1	
Other	2 (amantadine)	
Partner (Y/N)	21/5	
modified H–Y-stage[b]		
Stage 1	9	
Stage 1.5	4	
Stage 2	8	
Stage 2.5	3	
Stage 3	1	
Stage 4	0	
Stage 5	0	
Fallers[b,c]	7	
Freezers[b,d]	16	
UPDRS-score[a]	41.19 (44.15)	15–71
UPDRS III[a]	23.27 (9.64)	8–39
Retropulsion test score[a,e]		
0 = normal	22	
1 = retropulsion, but recovers unaided	2	
2 = absence of postural response; would fall if not caught by examiner	2	
3 = very unstable, tends to lose balance spontaneously	0	
4 = unable to stand without assistance	0	
BBS-score[a]	53.77 (1.99)	46–56
Functional Reach[a] (cm)	33.54 (7.36)	22–50
PG[a]	2.15 (2.15)	0–9

[a] Mean, SD standard deviation (between brackets).
[b] Number of patients.
[c] Posture and Gait, item 1: score ≥ 1.
[d] Freezing of Gait Questionnaire, item 3: score ≥ 2.
[e] UPDRS, item 30, M, male; F, female; Y, yes; N, no; H–Y, Hoehn and Yahr Scale.

reliability ranged from 0.74 for the UPDRS-III and the Fr test to 0.88 for the step frequency on the timed 10 m walk test (Table 2). Bland Altman plots are displayed in Fig. 1, with the dashed and bold line representing the mean difference score between the two assessments and the two dashed lines are representing the limits of agreement. The analysis showed homogenous distribution of differences for all tests. No significant systematic differences ($p < 0.5$) are found between the test and retest, except for the step frequency in the 10 m walk test. A mean systematic difference of 2.32 steps/min was observed ($p = 0.03$).

3.2. Responsiveness

Table 2 illustrates the SDDs for the five performance tests. Responsiveness ranged from three points on the BBS to 13 points on the UPDRS-III, reflecting a RCI of 5% for the BBS up to 11% for the UPDRS-III (Table 2).

Table 2
Intraclass correlation coefficients for inter-observer reliability and intra-observer reliability, smallest detectable difference and reliable change index ($N = 26$)

Test	Inter observer reliability	Intra observer reliability	SDD	RCI (%)
UPDRS	0.78*	0.84*	15	10
UPDRS-III	0.68*	0.74*	13	11
TGUG (s)	0.85*	0.88*	1.63	
10 m walk test[a]				
walking speed (m/s)	0.87*	0.81*	0.19	
step frequency (steps/min)	0.80*	0.88*	13	
BBS	0.74*	0.87*	2.84	5
Fr (cm)	0.64*	0.74*	11.5	

SDD, Smallest detectable difference; RCI, Reliable change index; UPDRS, Unified Parkinson Disease Ratio Scale; III, motor examination section; TGUG, Timed Get Up and Go; PG, Posture and Gait Score; BBS, Berg Balance Scale; Fr, Functional Reach. *$p < 0.01$.
[a] The 10 m walk test was performed at comfortable walking speed.

3.3. Feasibility

The duration of applying the test battery ranged from 20 to 30 min at the first assessment, whereas the duration of the second assessment ranged from 20 to 25 min.

4. Discussion

In the present study, tests were evaluated in terms of reliability and responsiveness by three independent observers in the home environment of PD patients. In general the test battery shows moderate to excellent inter-observer reliability (ICC = 0.64–0.87), and moderate to excellent intra-observer reliability (ICC = 0.74–0.88) according to the classification of ICC's of Fleiss [29]. Good reliability was obtained regardless of the fact that the tests could not be completely standardised, due to the different interiors of the subjects homes. Interestingly, despite the lack of observer experience with using the test battery in PD patients, inter and intra-observer reliability was moderate to excellent. These findings underpin the robustness of the tests included in this battery. In addition, the average time needed to apply the whole test battery was about 25 min, ensuring its feasibility.

Our findings on reliability are comparable with the literature with respect to tests such as the UPDRS [7,13,15], TGUG [9], 10 m walk test [11,12]. For the Fr test a higher ICC for intra-observer reliability was observed (ICC = 0.64) compared to the ICC (0.42) found by Smithson et al. [10] for PD patients without a history of falls. The ICC found by

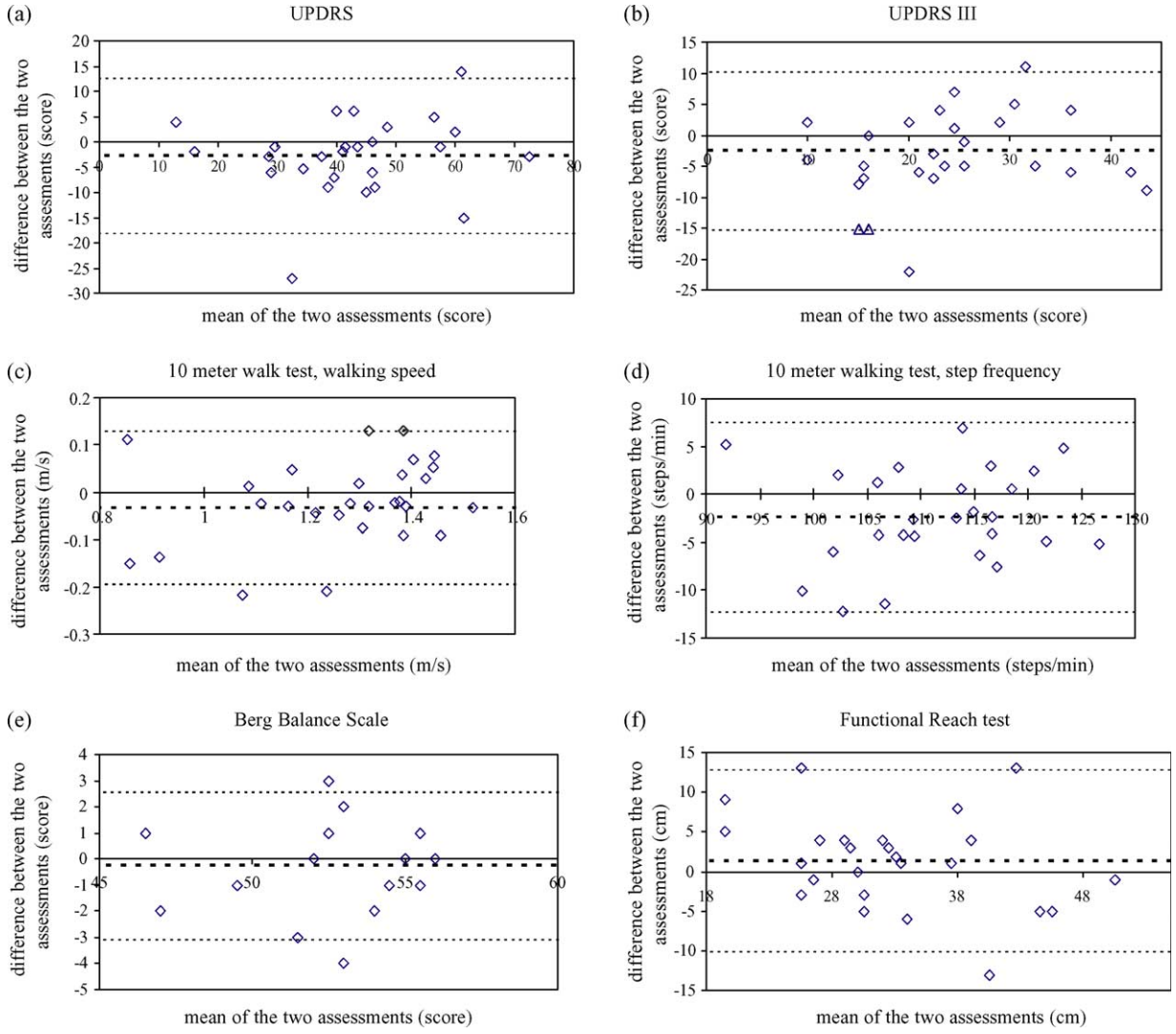

Fig. 1. Agreement of the 5 tests: graphic representation according to the Bland and Altman technique. UPDRS, Unified Parkinson Disease Ratio Scale; III, motor examination section. Dashed bold lines represent the mean difference score, dashed lines represent the limits of agreement, defined as the mean $\pm 1.96\times$ the standard deviation of the difference score

Smithson et al. [10] for PD patients with a history of falls was higher (ICC = 0.93) compared to the ICC found for non-fallers [10]. Although no explanation is given by Smithson and colleagues, differences in ICC's may be due to the larger between-subject variability in executing the Fr test in the population of fallers when compared to non-fallers. The PD patients in the current study were non-fallers ($n = 19$) and fallers ($n = 7$) which might have influenced the intra-observer reliability.

Due to the lack of consensus regarding methods of measuring responsiveness in the literature [30–32], SDDs in this study cannot directly be compared with earlier reported SDDs for the tests used in the current study. In most studies responsiveness was defined as the average changes in scores relative to baseline in self-rated clinically stable and improved patients [2,33]. Two studies were found that applied the same method for determining responsiveness on some of the tests used in the present study [34,35]. In these studies where the BBS and the timed 10 m walk test were applied, the SDDs for both tests were 6 points [34] and 0,16 m/s [35], respectively. However these tests were investigated in patients with stroke and therefore difficult to generalize to the PD patients in the present study. Unfortunately, general accepted criteria to judge the responsiveness of these tests do not exist. Therefore, in order to interpret the calculated scores, we compared the RCI's with each other. All RCI's were 11% or less, which is in our opinion acceptable for clinical use. In particular, realizing that assessments in patients own home environment is accompanied with a decreased ability to standardize the execution of measurements and with that higher error rates.

A limitation of the present study was that only subjects with mild to moderate disease severity were included (H and Y score ranging from 1 to 3). The battery was composed for those who were independent in ADL and were able to perform the tests without the use

of a walking aid. This limits the generalisation of present findings to the population of PD patients in general. In addition, all patients were tested in the on-phase and each patient was assessed at approximately the same time of the day, however, differences in time of assessment relative to moment of medication intake could have influenced the present findings.

5. Conclusion

The test battery was shown to have moderate to excellent reliability, despite limited clinical experience of the observers in assessing PD patients. Agreement on practical guidelines is necessary to standardize assessments as much as possible under less optimal circumstances, such as the home situation.

Since there appears to be a lack of consensus on how to quantify responsiveness [32], strict comparison with the literature is difficult. However, the results from the current study can be applied as indicators for an approximate threshold in the utility of the tests as outcome measures in a larger clinical trial.

Acknowledgements

This research project was supported by a grant from the European Commission (QLK6-CT-2001-00120; Rehabilitation in Parkinson's Disease: Strategies for Cueing).

References

[1] Morris ME. Movement disorders in people with Parkinson disease: a model for physical therapy. Phys Ther 2000;80(6):578–97.
[2] Goetz CG, Poewe W, Rascol O, Sampaio C, Glenn T, Stebbins GT. The unified Parkinson's disease rating scale (UPDRS): status and recommendations. Mov Disord 2003;18(7):738–50.
[3] Ramaker C, Marinus J, Stiggelbout AM, Van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 2002;17(5):867–76.
[4] Sampaio C, Ferreira JJ, Costa J. Evidence-based medicine (EBM) applied to Parkinson's disease treatment. Parkinsonism Relat Disord 2002;9(1):7–13.
[5] Rascol O, Goetz C, Koller W, Poewe W, Sampaio C. Treatment interventions for Parkinson's disease: an evidence based assessment. Lancet 2002;359(9317):1589–98.
[6] WHO. International classification of functioning, disability and health—ICF. Geneva: WHO; 2001.
[7] Richards M, Marder K, Cote L, Mayeux R. Interrater reliability of the unified Parkinson's disease rating scale motor examination. Mov Disord 1994;9(1):89–91.
[8] Stebbins GT, Goetz CG. Factor structure of the Unified Parkinson's Disease Rating Scale: Motor Examination section. Mov Disord 1998; 13(4):633–6.
[9] Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the timed up and go test in people with Parkinson disease. Phys Ther 2001;81(2):810–8.
[10] Smithson F, Morris ME, Iansek R. Performance on clinical tests of balance in Parkinson's disease. Phys Ther 1998;78(6):577–92.
[11] Urquhart DM, Morris ME, Iansek R. Gait consistency over a 7-day interval in people with Parkinson's disease. Arch Phys Med Rehabil 1999;80(6):696–701.
[12] Morris ME, Matyas TA, Iansek R, Summers JJ. Temporal stability of gait in Parkinson's disease. Phys Ther 1996;76(7):763–77.
[13] Bennett DA, Shannon KM, Beckett LA, Goetz CG, Wilson RS. Metric properties of nurses ratings of parkinsonian signs with a modified unified Parkinson's disease rating scale. Neurology 1997;49(6): 1580–7.
[14] Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries J, Bermejo F. Unified Parkinson's disease rating scale characteristics and structure. The Cooperative Multicentric Group. Mov Disord 1994;9(1):76–83.
[15] Siderowf A, McDermott M, Kieburtz K, Blindauer K, Plumb S, Shoulson I. Test-retest reliability of the unified Parkinson's disease rating scale in patients with early Parkinson's disease: results from a multicenter clinical trial. Mov Disord 2002;17(4):758–63.
[16] Nieuwboer A, De Weerdt W, Dom R, Truyen M, Janssens L, Kamsma Y. The effect of a home physiotherapy program for persons with Parkinson's disease. J Rehabil Med 2001;33(6):266–72.
[17] Nieuwboer A, De Weerdt W, Dom R, Bogaerts K. Prediction of outcome of physiotherapy in advanced Parkinson's disease. Clin Rehabil 2002;16(8):886–93.
[18] Jones D, Rochester L, Hetherington V, Baker K, Nieuwboer A, Willems A-M, Chavret F, Kwakkel G, Van Wegen E, Lim I. The Rescue project: developing and testing an optimized cueing rehabilitation programme for people with Parkinson's disease Proceeding from: Multidisciplinary Care in Parkinson's Disease and Parkinsonism- From Science to Practice 2003.
[19] Hoehn MM, Yahr MD. Parkinsonism: onset, progression, and mortality, 1967. Neurology 2001;57(10 Suppl 3):S11–S26.
[20] Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189–98.
[21] Podsiadlo D, Richardson S. The timed up and go: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991; 39(2):142–8.
[22] Berg KO, Maki BE, Williams JI, Holliday PJ, Wood-Dauphinee SL. Clinical and laboratory measures of postural balance in an elderly population. Arch Phys Med Rehabil 1992;73(11):1073–80.
[23] Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. J Gerontol 1990;45(6):M192–M197.
[24] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420–8.
[25] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476): 307–10.
[26] Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 1998; 12(3):187–99.
[27] Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001;10(7):571–8.
[28] Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003; 56(5):395–407.
[29] Fleiss JL. The design and analysis of clinical experiments 1986. New York, Wiley.
[30] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53(5):459–68.
[31] Wallace D, Duncan PW, Lai SM. Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. J Clin Epidemiol 2002;55(9): 922–8.

[32] van Bennekom CA, Jelles F, Lankhorst GJ, Bouter LM. Responsiveness of the rehabilitation activities profile and the Barthel index. J Clin Epidemiol 1996;49(1):39–44.
[33] Peto V, Jenkinson C, Fitzpatrick R. Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. Age Ageing 2001;30(4):299–302.
[34] Stevenson TJ. Detecting change in patients with stroke using the Berg balance scale. Aust J Physiother 2001;47(1):29–38.
[35] Kwakkel G, Kollen BJ, Wagenaar RC. Long term effects of intensity of upper and lower limb training after stroke: a randomised trial. J Neurol Neurosurg Psychiatry 2002;72(4): 473–9.