

ORIGINAL REPORT

EXCELLENT CROSS-CULTURAL VALIDITY, INTRA-TEST RELIABILITY AND CONSTRUCT VALIDITY OF THE DUTCH RIVERMEAD MOBILITY INDEX IN PATIENTS AFTER STROKE UNDERGOING REHABILITATION

Leo D. Roorda, MD, PT, PhD¹, John Green, PT, PhD², Kiki R. A. De Kluis, MSc, MD³, Ivo W. Molenaar, PhD⁴, Pam Bagley, PT, PhD⁵, Jane Smith, RN, MSc² and Alexander C. H. Geurts, MD, PhD^{2,6}

From the ¹Department of Rehabilitation Medicine and Psychology, Jan van Breemen Institute, Amsterdam, The Netherlands, ²Academic Unit of Elderly Care and Rehabilitation, University of Leeds, Bradford, UK, ³Sint Maartens-kliniek, Nijmegen, ⁴Department of Statistics and Measurement, University of Groningen, Groningen, The Netherlands, ⁵Division of Rehabilitation, School of Health Studies, University of Bradford, Bradford, UK and ⁶Department of Rehabilitation, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

Objective: To investigate the cross-cultural validity of international Dutch-English comparisons when using the Dutch Rivermead Mobility Index (RMI), and the intra-test reliability and construct validity of the Dutch RMI.

Methods: Cross-cultural validity was studied in a combined data-set of Dutch and English patients undergoing rehabilitation after stroke, who were assessed with the Dutch version of the RMI and the original English RMI, respectively. Mokken scale analysis was used to investigate unidimensionality, monotone homogeneity model fit, and differential item functioning between the Dutch and the English RMI. Intra-test reliability and construct validity were studied in the Dutch patients by calculating the reliability coefficient and correlating the Dutch RMI and the Dutch Barthel Index.

Results: The RMI was completed for Dutch ($n=200$) and English ($n=420$) patients after stroke. The unidimensionality and monotone homogeneity model fit of the RMI were excellent: combined Dutch-English data-set (coefficient $H=0.91$); Dutch data-set (coefficient $H=0.93$); English data-set (coefficient $H=0.89$). No differential item functioning was found between the Dutch and the English RMI. The intra-test reliability of the Dutch RMI was excellent (coefficient $\rho=0.97$). In a sub-sample of patients ($n=91$), the Dutch RMI correlated strongly with the Dutch Barthel Index (Spearman's correlation coefficient $\rho=0.84$).

Conclusion: The Dutch RMI allows valid international Dutch-English comparisons, and has excellent intra-test reliability and construct validity.

Key words: cerebrovascular accident, quality of life, disability evaluation, psychometrics.

J Rehabil Med 2008; 40: 727–732

Correspondence address: Leo D. Roorda, Department of Rehabilitation Medicine and Psychology, Jan van Breemen Institute, Dr. J. van Breemenstraat 2, 1056 AB Amsterdam, The Netherlands. E-mail: l.roorda@janvanbreemen.nl

Submitted September 18, 2007; accepted May 19, 2008

INTRODUCTION

The past 25 years have seen the development of a large number of patient-reported outcome (PRO) instruments, and these instruments are being used increasingly in scientific research, with a growing emphasis on international applications (1). International applications stem from the need to pool data across the nationality of respondents (e.g. for systematic reviews) and from increasing international research collaboration (2, 3).

The cross-national use of measurement instruments has created a need for cross-culturally valid instruments for outcome assessment (3). Readers of international journals may assume that the effects of an intervention that has been applied in another country can be generalized to their own situation if the setting, the study population and the intervention at hand are comparable. For instance, they may assume that, if exercise therapy increases mobility by 5 points in Dutch patients after stroke (measured with the Dutch Rivermead Mobility Index (RMI)), exercise therapy will also increase mobility by 5 points in English patients after stroke (measured with the English RMI). However, this assumption is based on the premise that the response of Dutch and English patients after stroke to the Dutch RMI and the English RMI, respectively, will be equivalent, and will depend on the patients' level of mobility, and not on their nationality; furthermore, any changes in mobility will result in equivalent changes in the scores for these different instruments. This issue of the comparability of scores between different cultural (including language) groups can be addressed in cross-cultural validation studies (3).

The RMI (4) is a PRO instrument that measures mobility, an important aspect of daily functioning in patients after stroke, and is being used increasingly for international research in patients with stroke. The clinimetric properties of the original English version of the RMI are well known (4–7), and an increasing number of translations are available (8–10). However, as yet, no cross-cultural validation study has been carried out. From this perspective, we have recently made a Dutch translation of the RMI.

The purpose of this study was to investigate the cross-cultural validity, the intra-test reliability and the construct validity of this Dutch RMI in patients undergoing rehabilitation after stroke.

METHODS

Patients

We recruited a cohort of patients in the Netherlands (further referred to as Dutch patients) with a definite diagnosis of stroke who had been admitted to a specialized department of a rehabilitation centre (Sint Maartenskliniek in Nijmegen). Two-hundred (94%) of the 212 patients after stroke who were admitted to the department between January 2001 and January 2005 participated in the study. Twelve patients did not participate, mainly for logistic reasons due to the short length of their stay.

Data-sets from 3 different English studies were used to study cross-cultural validity. The first study (11) concerned English patients living in the community with mobility problems more than one year after stroke, who participated in a randomized controlled trial assessing the effectiveness of physiotherapy. Of the potential 359 trial participants 182 fulfilled the selection criteria, and 171 (93%) completed a baseline assessment. Data on these 171 patients were used for the current study. The second study (12) concerned English patients with a clinical diagnosis of stroke, admitted to an inpatient stroke rehabilitation unit, who participated in a randomized controlled trial assessing the effectiveness of the Oswestry standing frame. Of the 412 patients referred to the stroke unit, 167 fulfilled the inclusion criteria and 140 (84%) participated in the study. Data on these 140 participants were used in the current study. The third study (13) concerned English patients admitted to a stroke rehabilitation unit. Of the 122 potential participants, 109 (89%) completed an assessment 3 weeks after stroke onset, and these data were used in the current study. Thus, the total number of English patients who fulfilled the inclusion criteria was 471, and the total number of participants was 420 (89%).

Measurements

Sociodemographic and clinical characteristics were obtained from medical records. The Dutch patients were assessed directly after admission to the rehabilitation centre. The physical therapists involved in the treatment of these patients completed the Dutch RMI for all participants, and the Dutch Barthel Index (BI) for a sub-sample of the participants. With respect to the first and second English studies (11, 12) we used data from the baseline assessments, which were collected by research physiotherapists. The RMI of some participants could not be obtained at baseline in the acute phase after stroke in the third English study (13). We therefore used the RMI data from the 3-week follow-up assessment, which were collected by a research nurse.

The original English RMI is a PRO measure of mobility (Appendix 1) (4). It is simple to use, clinically relevant, and has been well-tested, with satisfactory reliability, validity and responsiveness (4–10, 14–16). The RMI is an ordinal scale, consisting of 15 items addressing several aspects of mobility, and each item has 2 response options: yes or no. The sum scores range from 0 to 15, with higher scores indicating better mobility. The Dutch version of the RMI was translated by 2 bilingual and bicultural translators who were familiar with patients after stroke, in a double (back-) translation procedure (Appendix 2) (17). In order to make the RMI fully applicable in daily clinical practice (to be able to include patients with communication and cognitive problems) we slightly adapted the RMI instructions in the translation process: in the English RMI, except for one observer-based question, the assessor is instructed to ask the patient the questions (patient-reported), whereas in the Dutch RMI the assessor is instructed to observe the patient (observer-based).

The BI is a measure of mobility and personal care (18, 19), which is widely used and has been well-tested, and has been found to have satisfactory reliability, validity and responsiveness (7, 15, 19–21). The

BI consists of 10 items, each of which has 2–4 possible responses. The sum scores for the BI range from 0 to 20, with higher values indicating better mobility and personal care.

Statistics

Patients. Differences in sociodemographic and clinical characteristics between the Dutch and the English patients were investigated with an independent samples *t*-test (age) and χ^2 tests (gender and stroke location).

Unidimensionality and monotone homogeneity model fit. Unidimensionality indicates that the items (questions) of a measurement instrument assess one single underlying construct (22), whereas the fit of the items with the monotone homogeneity (MH) model implies that the items can be used for the measurement of patients (23–26).

We used Mokken scale analysis to investigate both unidimensionality and fit with the MH model of the RMI items (23–25). Mokken scale analysis can be considered as a non-parametric approach to item response theory (IRT). IRT is the class of psychometric models for scale construction that assumes that observed responses (answers) to items can be explained by a latent trait (variable), in our case mobility. Mokken scale analysis provides ordinal information about the location of patients and items on the scale of the latent trait. Patients are ordered on this scale according to their sum scores: patients with higher sum scores have better mobility. Items are ordered on the scale of the latent trait according to their mean score, which is the proportion of patients who respond positively to the item. A positive response to an item with a low mean score indicates better mobility.

Within the framework of Mokken scale analysis, unidimensionality can be studied with the SEARCH procedure in the MSP software (24, 25). We investigated unidimensionality by stepwise increasing *c*, which is the lower bound for the scalability coefficient *H* (see below). According to Hemker et al. (27), a multidimensional item bank will often appear to form one scale at *c*=0.30, while an item bank consisting of unidimensional scales will directly split into its unidimensional scales. At intermediate values of *c* (between 0.40 and 0.60) a multidimensional scale will often break up into subscales. Beyond values of 0.80, unidimensional scales will also tend to split into their individual items.

Within the framework of Mokken scale analysis, the fit of the MH model is evaluated with the TEST procedure in the MSP software (24, 25). MH model fit is evaluated by calculating the scalability coefficient *H*, which is a global indicator of the degree to which patients can be accurately ordered on the latent trait by means of their sum score. Scale criteria are met when: (i) the coefficients of scalability for all item pairs (*H_{ij}*) are positive; (ii) the scalability coefficients for the items in relation to the scale at issue (*H_i*) are at least 0.30; and (iii) the scalability coefficient for the scale (*H*) is at least 0.30. Higher values for *H_i* and *H* imply a better scale. A rule of thumb is that a scale is considered to be strong when $H \geq 0.50$, medium when $0.50 > H \geq 0.40$, and weak when $0.40 > H \geq 0.30$ (24, 25).

Differential item functioning. Differential item functioning (DIF), or item bias, addresses the issue of making valid comparisons between subgroups of patients. One may consider 2 subgroups of patients, for example Dutch and English patients after stroke, and assume that a certain item functions differently in these patients. In such a case, Dutch patients with the same true mobility as English patients may more often respond positively to this item, and they will then unintentionally tend to have higher scores than the English patients. So, an item that functions differently in subgroups of patients causes differences in subgroup scores, even when the patients in the subgroups have similar mobility. As a consequence, DIF impedes valid comparisons between these subgroups.

In this study we investigated DIF between the Dutch and the English RMI. Within the framework of Mokken scale analysis, DIF is studied by checking the assumption of equal ordering of the items on the scale of the latent trait. DIF is found if the ordering of the items is different within the subgroups that are investigated. For a detailed

check of DIF a diagnostic *Crit* value is calculated (24). No DIF is found if the largest *Crit* value per item is less than 40, but if the *Crit* value exceeds 80, DIF is suggested. DIF results can also be presented in the form of a scatter plot.

Intra-test reliability. Intra-test reliability (or internal consistency) assesses the degree of repeatability of the sum score. We quantified the intra-test reliability by calculating the reliability coefficient ρ (24, 25). This is slightly superior to Cronbach's α , which underestimates intra-test reliability when there is substantial variation in the level of mean scores per item (28). A reliability coefficient of 0.90 or more is recommended for stable decisions about individual patients (22).

Construct validity. Construct validity assesses the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses for the constructs that are being measured (22). We expected that the sum scores of the Dutch RMI and the Dutch BI would be strongly and positively correlated. Spearman's correlation coefficient ρ was calculated to evaluate construct validity.

RESULTS

Patients

The mean (standard deviation (SD)) age of the Dutch patients was 58.6 (11.8) years; 102 (51%) were male. A supratentorial stroke had been sustained by 66 patients in the right hemisphere, and 91 patients in the left hemisphere. Forty-three patients had sustained an infra-tentorial stroke. The mean (SD) age of the English patients was 74.3 (9.2) years; 195 (46%)

were male. A supratentorial stroke had been sustained by 203 patients in the right hemisphere, and 213 patients in the left hemisphere. Four patients had sustained an infra-tentorial stroke. There were significant differences between the Dutch and the English patients in age (2-tailed $p < 0.000$) and location of stroke (2-tailed $p < 0.000$), but not in gender (2-tailed $p = 0.29$).

Unidimensionality

All the items in the combined Dutch-English data-set fitted in the scale, even at $c = 0.80$. This was also the case for the Dutch and English data-sets when they were tested separately. This indicates that the combined RMI, and the Dutch and the English RMI, are all strong unidimensional scales.

Monotone homogeneity model fit

Scale criteria were met for all items in the combined Dutch-English data-set and for both the Dutch and the English data-sets when they were tested separately (Table I). For the 15 items in each of these 3 data-sets the coefficients of scalability for the item pairs (H_{ij}) were all positive, the scalability coefficients for the items in relation to the scale at issue (H_i) were all greater than 0.30, and the scalability coefficients of the scales (H) were 0.91 for the combined Dutch-English data-set, 0.93 for the Dutch data-set, and 0.89 for the English data-set, indicating

Table I. Item mean scores, coefficients of scalability of the items and the scale, coefficient of reliability and median (interquartile range) sum scores of the combined Dutch-English, Dutch and English Rivermead Mobility Index

Item No.	Abbreviated text*	Combined Dutch-English (n=620)		Dutch (n=200)		English (n=420)	
		Mean†	H_i ‡	Mean	H_i	Mean	H_i
15.	Running	0.04	1.00	0.12	1.00	0.00	§
14.	Up and down four steps	0.09	0.96	0.21	0.97	0.03	0.92
12.	Walking outside (uneven ground)	0.16	0.93	0.24	0.98	0.12	0.89
13.	Bathing	0.20	0.87	0.32	0.90	0.14	0.84
9.	Walking outside (even ground)	0.26	0.89	0.33	0.95	0.23	0.85
8.	Stairs	0.27	0.87	0.37	0.91	0.23	0.84
10.	Walking inside, with no aid	0.29	0.90	0.35	0.94	0.25	0.88
11.	Picking off floor	0.31	0.92	0.34	0.94	0.29	0.90
7.	Walking inside, with an aid if needed	0.44	0.95	0.46	0.97	0.43	0.94
6.	Transfer	0.51	0.95	0.63	0.93	0.45	0.95
5.	Standing unsupported	0.55	0.93	0.69	0.90	0.48	0.94
4.	Sitting to standing	0.56	0.94	0.72	0.93	0.49	0.94
2.	Lying to sitting	0.61	0.89	0.76	0.89	0.54	0.88
1.	Turning over in bed	0.75	0.81	0.90	0.88	0.67	0.77
3.	Sitting balance	0.75	0.86	0.93	0.92	0.67	0.84
Scalability coefficient H		0.91		0.93		0.89	
Reliability coefficient ρ #		0.97		0.97		0.96	
Sum score median**		5		6		3	
Sum score interquartile range		1-10		3-12		1-9	

*Abbreviated item text. The non-abbreviated item text can be found in Appendices 1 and 2.
 †Item mean score, indicating the proportion of patients responding positively to the items. A positive response to an item with a low mean score indicates better mobility.
 ‡Scalability coefficient H_i of the items in relation to the scale (range 0-1 under the monotone geneity (MH) model). A minimum value of $H_i = 0.30$ is recommended.
 §Scalability coefficient H_i could not be obtained because none of the English patients responded positively to this item.
 || Scalability coefficient H for the scale (range 0-1 under the MH model). A scale is considered to be strong when $H \geq 0.50$.
 #Reliability coefficient ρ for the scale (range 0-1 under the MH model). $\rho \geq 0.90$ is recommended for decisions about individual patients.
 **Sum scores range from 0 to 15. Patients with higher sum scores have better mobility.

strong scales. These results demonstrate that the RMI items can be used for the measurement of patients.

Differential item functioning

All *Crit* values were less than 40, and no DIF was found between the items of the Dutch and the English RMI (Fig. 1). This indicates that valid comparisons can be made between Dutch patients after stroke responding to the Dutch RMI and English patients responding to the English RMI.

Intra-test reliability

The intra-test reliability coefficient ρ was 0.97. This indicates that the intra-test reliability of the Dutch RMI was excellent, both for group descriptions and for decisions about individual patients.

Construct validity

The Dutch RMI sum score (median (interquartile range (IQR)) 7 (4–13)) was strongly and positively correlated (Spearman’s correlation coefficient $\rho=0.84$) with the Dutch BI sum score (median (IQR) 12 (9–17)) in a sub-sample of 91 patients. This indicates very good construct validity.

DISCUSSION

We investigated the cross-cultural validity of international Dutch-English comparisons when applying the Dutch RMI

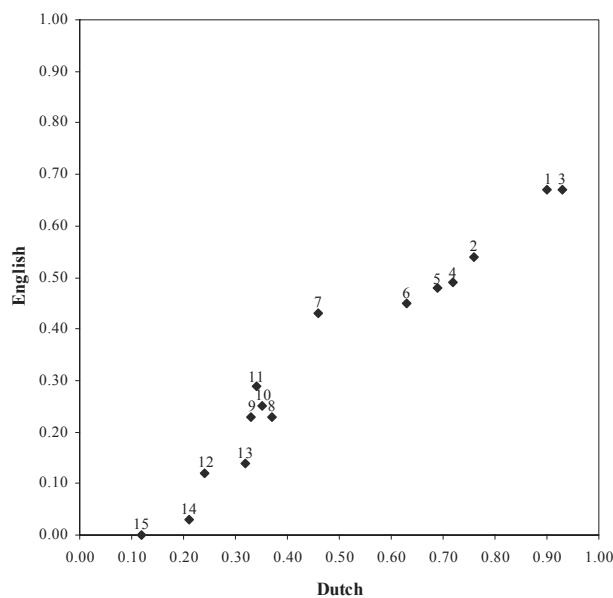


Fig. 1. Mean item scores for Dutch patients after stroke responding to the Dutch Rivermead Mobility Index (RMI), and English patients after stroke responding to the English RMI. Mean scores for the Dutch and English patients are presented on the x and y axes, respectively. A positive response to an item with a low mean score indicates better mobility. The number above each data-point refers to the RMI item number (Table I). Items demonstrating differential item functioning would show major deviations from an imaginary line that can be drawn through the data-points.

to patients after stroke. There were significant differences between the Dutch and the English patients after stroke with respect to age and stroke location. However, separate analyses (data not presented) demonstrated that age and stroke location resulted in no difference in the order of the items, which implies that, despite these significant differences in age and stroke location, the cross-cultural validity of the Dutch RMI could be satisfactorily investigated in the current Dutch and English data-sets.

We studied unidimensionality and the item parameters of the Dutch and the English RMI with a Mokken scale analysis, which is a non-parametric IRT method. In a similar study addressing the cross-cultural validity of the Western Ontario and McMaster Universities osteoarthritis index we used Rasch analysis (29), which is a more frequently used parametric IRT method (30–33). However, the advantages of a non-parametric IRT model are that response data will more easily fit with such a model and that the sum score is more easily understood. Moreover, because the RMI is an ordinal scale, we think that a non-parametric IRT model is more appropriate for this instrument.

In this study we demonstrated unidimensionality and an excellent fit with the MH model for both the Dutch and the English RMI, which implies that both versions do indeed measure a single construct (mobility), and that both are suitable for measuring patients after stroke. As far as we are aware, fit with the MH model has not previously been investigated for the RMI.

In order to study cross-cultural validity we used data-sets from 3 different English studies. One advantage of combining data-sets is that the scale analysis will result in better estimates of the scale parameters because of the larger number of patients included in the analysis. Moreover, in the current study, combining the 3 data-sets resulted in a larger variation in patient mobility levels. A disadvantage might be that 3 "different" scales are combined. In an additional analysis (data not shown) we found excellent fit of the combined English RMI item set with the double monotonicity (DM) model from Mokken scale analysis (24, 25). Examples of studies addressing the DM fit of an item set are available from the literature (34, 35). Fit with the DM model indicates that the (hierarchical) ordering of the RMI items is the same for patients with different levels of mobility. Thus, combining different data-sets, even of patients with different levels of mobility, will not influence the ordering of the items. This implies that the cross-cultural validity of the Dutch RMI could be investigated in the 3 combined English data-sets.

We found no DIF between the Dutch RMI and the English RMI, which indicates that valid comparisons can be made between Dutch and English patients after stroke. Applying these results to the current data-sets, it can be shown that the Dutch patients after stroke had statistically significant higher RMI scores than the English patients (median (IQR): 6 (3–12) vs 3 (1–9); Mann-Whitney *U* test, 2-tailed $p<0.000$) (Table I), and this may be related to the age difference between the 2 groups. Our results enable researchers to pool data-sets, which were obtained with either the Dutch or the English

RMI. In addition, readers of international journals can make a straightforward comparison of scores on the Dutch and the English RMI. It should be mentioned here that changing the RMI from a patient-reported (English) into an observer-based (Dutch) instrument did not influence the ordering of the items in the Dutch and English samples that we studied.

The results of this study also demonstrated excellent intra-test reliability and construct validity of the Dutch RMI. However, we only studied intra-test reliability. Further studies should therefore address other aspects of reliability, such as intra-rater (or test-retest) reliability and inter-rater reliability. Furthermore, with regard to construct validity, we only studied the correlation of the Dutch RMI with the Dutch BI. Construct validity should therefore be studied in more detail in future research. Finally, the responsiveness (or sensitivity to change) (36) of the Dutch RMI should also be studied in future research. We would, however, expect satisfactory results with respect to the clinimetric properties (reliability, validity and responsiveness), which have not been addressed in the current study, because, firstly, the original English RMI has good reliability, validity and responsiveness (4–10, 14–16) and, secondly, the results of the current study indicate the absence of DIF between the Dutch and the English RMI.

In summary, the Dutch RMI allows valid international Dutch-English comparisons, and has excellent intra-test reliability and construct validity.

REFERENCES

- Acquadro C, Jambon B, Ellis D, Marquis P. Language and translation issues. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 575–585.
- Anderson RT, Aaronson NK, Leplege AP, Wilkin D. International use and application of generic health-related quality of life instruments. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 613–632.
- Bullinger M, Power MJ, Aaronson NK, Cella DF, Anderson RT. Creating and evaluating cross-cultural instruments. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 659–668.
- Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. *Int Disabil Stud* 1991; 13: 50–54.
- Forlander DA, Bohannon RW. Rivermead Mobility Index: a brief review of research to date. *Clin Rehabil* 1999; 13: 97–100.
- Rossier P, Wade DT. Validity and reliability comparison of 4 mobility measures in patients presenting with neurologic impairment. *Arch Phys Med Rehabil* 2001; 82: 9–13.
- Green J, Forster A, Young J. A test-retest reliability study of the Barthel Index, the Rivermead Mobility Index, the Nottingham Extended Activities of Daily Living Scale and the Frenchay Activities Index in stroke patients. *Disabil Rehabil* 2001; 23: 670–676.
- Hsieh CL, Hsueh IP, Mao HF. Validity and responsiveness of the rivermead mobility index in stroke patients. *Scand J Rehabil Med* 2000; 32: 140–142.
- Schindl MR, Forstner C, Kern H, Zipko HT, Rupp M, Zifko UA. Evaluation of a German version of the Rivermead Mobility Index (RMI) in acute and chronic stroke patients. *Eur J Neurol* 2000; 7: 523–528.
- Franchignoni F, Tesio L, Benevolo E, Ottonello M. Psychometric properties of the Rivermead Mobility Index in Italian stroke rehabilitation inpatients. *Clin Rehabil* 2003; 17: 273–282.
- Green J, Forster A, Bogle S, Young J. Physiotherapy for patients with mobility problems more than 1 year after stroke: a randomised controlled trial. *Lancet* 2002; 359: 199–203.
- Bagley P, Hudson M, Forster A, Smith J, Young J. A randomized trial evaluation of the Oswestry Standing Frame for patients after stroke. *Clin Rehabil* 2005; 19: 354–364.
- Smith J, Brotheridge S, Young J. Patterns of hemiparesis recovery in lacunar and partial anterior circulation infarct stroke syndromes. *Clin Rehabil* 2001; 15: 59–66.
- Antonucci G, Aprile T, Paolucci S. Rasch analysis of the Rivermead Mobility Index: a study using mobility measures of first-stroke inpatients. *Arch Phys Med Rehabil* 2002; 83: 1442–1449.
- Hsueh IP, Wang CH, Sheu CF, Hsieh CL. Comparison of psychometric properties of three mobility measures for patients with stroke. *Stroke* 2003; 34: 1741–1745.
- Chen HM, Hsieh CL, Sing KL, Liaw LJ, Chen SM, Lin JH. The test-retest reliability of 2 mobility performance tests in patients with chronic stroke. *Neurorehabil Neural Repair* 2007; 21: 347–352.
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25: 3186–3191.
- Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965; 14: 61–65.
- Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. *Int Disabil Stud* 1988; 10: 61–63.
- 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: 922–928.
- Sainsbury A, Seebass G, Bansal A, Young JB. Reliability of the Barthel Index when used with older people. *Age Ageing* 2005; 34: 228–232.
- Nunnally JC, Bernstein IH, editors. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
- Mokken RJ. Nonparametric models for dichotomous responses. In: Hambleton RK, editor. *Handbook of modern item response theory*. New York: Springer; 1997, p. 351–367.
- Molenaar IW, Sijtsma K. *User's manual MSP5 for windows*. Groningen: iecProGAMMA; 2000.
- Sijtsma K, Molenaar IW, editors. *Introduction to nonparametric item response theory*. Thousand Oaks: Sage; 2002.
- Sijtsma K, Emons WH, Bouwmeester S, Nyklicek I, Roorda LD. Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Qual Life Res* 2008; 17: 275–290.
- Hemker BT, Sijtsma K, Molenaar IW. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Appl Psychol Measurement* 1995; 19: 337–352.
- Mokken RJ, editor. *A theory and procedure of scale analysis with applications in political research*. The Hague: Mouton; 1971.
- Roorda LD, Jones CA, Waltz M, Lankhorst GJ, Bouter LM, van der Eijken JW, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004; 63: 36–42.
- Andrich D, editor. *Rasch models for measurement*. Newbury Park: Sage; 1988.
- Cella DF, Lloyd SR, Wright BD. Cross-cultural instrument equating: current research and future directions. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 707–715.
- Wright BD, Stone MH, editors. *Best test design*. Chicago: MESA; 1979.
- Wright BD, Masters GN, editors. *Rating scale analysis*. Chicago: MESA; 1982.

34. Roorda LD, Roebroek ME, van Tilburg T, Lankhorst GJ, Bouter LM. Measuring activity limitations in climbing stairs: development of a hierarchical scale for patients with lower-extremity disorders living at home. *Arch Phys Med Rehabil* 2004; 85: 967–971.
35. Roorda LD, Roebroek ME, van Tilburg T, Molenaar IW, Lankhorst GJ, Bouter LM, et al. Measuring activity limitations in walking: development of a hierarchical scale for patients with lower-extremity disorders who live at home. *Arch Phys Med Rehabil* 2005; 86: 2277–2283.
36. Juniper EF, Guyatt GH, Jeaschke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2nd edn. Philadelphia: Lippincott-Raven; 1996, p. 49–56.

APPENDIX 1. English version of the Rivermead Mobility Index

Instructions

The patient is asked the following 15 questions, but item 5 is observed. A score of 1 is given for each “yes” answer.

1. Turning over in bed
Do you turn over from your back to your side without help?
2. Lying to sitting
From lying in bed, do you get up to sit on the edge of bed on your own?
3. Sitting balance
Do you sit on the edge of the bed without holding on for 10 sec?
4. Sitting to standing
Do you stand up (from any chair) in less than 15 sec, and stand there for 15 sec (using hands, and with an aid if necessary)?
5. Standing unsupported
Observe standing for 10 sec without any aid.
6. Transfer
Do you manage to move from bed to chair and back without any help?
7. Walking inside, with an aid if needed
Do you walk 10 m, with an aid if necessary, but with no standby help?
8. Stairs
Do you manage a flight of stairs without help?
9. Walking outside (even ground)
Do you walk around outside, on pavements without help?
10. Walking inside, with no aid
Do you walk 10 m inside with no calliper, splint, or aid, and no standby help?
11. Picking off floor
If you drop something on the floor, do you manage to walk 5 m, pick it up and then walk back?
12. Walking outside (uneven ground)
Do you walk over uneven ground (grass, gravel, dirt, snow, ice, etc.) without help?
13. Bathing
Do you get in/out of bath or shower unsupervised and wash self?
14. Up and down four steps
Do you manage to go up and down four steps with no rail and without help, but using an aid if necessary?
15. Running
Do you run 10 m without limping in 4 sec (fast walk is acceptable)?

APPENDIX 2. Dutch version of the Rivermead Mobility Index

Instructie

De therapeut observeert de patiënt bij het uitvoeren van onderstaande activiteiten. Score 1 wordt gegeven voor iedere activiteit die zelfstandig kan worden uitgevoerd.

1. Omrollen in bed
Kan de patiënt in bed vanuit ruglig naar de zij rollen zonder hulp?
2. Van lig naar zit
Kan, de patiënt, als hij/zij in bed ligt, zonder hulp op de rand van het bed komen zitten?
3. Zitbalans
Kan de patiënt 10 tellen zonder steun of vast te houden op de rand van het bed zitten?
4. Van zit naar stand
Kan de patiënt vanuit een stoel binnen 15 sec komen staan en 15 sec blijven staan (met gebruik van handen en/of hulpmiddel indien nodig)?
5. Stabalans
Kan de patiënt staan zonder steun gedurende 10 sec?
6. Transfer
Kan de patiënt van het bed naar de stoel komen en terug zonder hulp?
7. Lopen in huis, met hulpmiddel indien nodig
Kan de patiënt zelfstandig 10 m lopen, met hulpmiddel indien nodig?
8. Traplopen
Kan de patiënt zelfstandig de trap op en af lopen?
9. Lopen buiten, op effen terrein
Kan de patiënt buiten lopen, op het trottoir zonder hulp (eventueel met hulpmiddel)?
10. Lopen binnen zonder hulpmiddel
Kan de patiënt zelfstandig in huis lopen zonder hulpmiddel of orthese?
11. Iets oppakken van de grond
Als de patiënt iets op de grond laat vallen, kan hij/zij dan 5 m lopen, het voorwerp oppakken en weer terug lopen?
12. Lopen buiten, op oneffen terrein
Kan de patiënt buiten lopen op oneffen terrein (gras, grind, sneeuw, hellingen, stoepranden etc.) zonder hulp?
13. Baden/douchen
Kan de patiënt in en uit het bad/de douche komen en zichzelf wassen zonder hulp?
14. Vier treden op en af
Kan de patiënt vier treden op en af lopen zonder leuning, eventueel met gebruik van een loophulpmiddel?
15. Hardlopen
Kan de patiënt 10 m hardlopen binnen 4 sec (snelwandelen is toegestaan) in gelijk tred?