VU Research Portal

Measuring activity limitations in climbing stairs

Roorda, L.D.; Roebroek, M.E.; van Tilburg, T.G.; Lankhorst, G.J.; Bouter, L.M.

published in

Archives of Physical Medicine and Rehabilitation
2004

DOI (link to publisher)

[10.1016/j.apmr.2003.11.018](https://doi.org/10.1016/j.apmr.2003.11.018)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Roorda, L. D., Roebroek, M. E., van Tilburg, T. G., Lankhorst, G. J., & Bouter, L. M. (2004). Measuring activity limitations in climbing stairs: development of a hierarchical scale for patients with lower-extremity disorders living at home. *Archives of Physical Medicine and Rehabilitation*, 85(6), 967-971.
<https://doi.org/10.1016/j.apmr.2003.11.018>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Measuring Activity Limitations in Climbing Stairs: Development of a Hierarchical Scale for Patients With Lower-Extremity Disorders Living at Home

Leo D. Roorda, MD, PT, Marij E. Roebroek, PhD, Theo van Tilburg, PhD, Gustaaf J. Lankhorst, MD, PhD, Lex M. Bouter, PhD, and the Measuring Mobility Study Group

ABSTRACT. Roorda LD, Roebroek ME, van Tilburg T, Lankhorst GJ, Bouter LM, Measuring Mobility Study Group. 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;85:967-71.

Objective: To develop a hierarchical scale that measures activity limitations in climbing stairs in patients with lower-extremity disorders living at home.

Design: Cross-sectional study with Mokken scale analysis of 15 dichotomous items.

Setting: Outpatient clinics of secondary and tertiary care centers.

Participants: Patients (N=759; mean age $\pm$ standard deviation, 59.8 $\pm$ 15.0y; 48% men) living at home, with different lower-extremity disorders: stroke, poliomyelitis, osteoarthritis, amputation, complex regional pain syndrome type I, and diabetic foot problems.

Interventions: Not applicable.

Main Outcome Measures: (1) Fit of the monotone homogeneity model, indicating whether items can be used for measuring patients; (2) fit of the double monotonicity model, indicating invariant (hierarchical) item ordering; (3) intratest reliability, indicating repeatability of the sum score; and (4) differential item functioning, addressing the validity of comparisons between subgroups of patients.

Results: There was (1) good fit of the monotone homogeneity model (coefficient $H=.50$) for all items for all patients, and for subgroups defined by age, gender, and diagnosis; (2) good fit of the double monotonicity model (coefficient $H^T=.58$); (3) good intratest reliability (coefficient $\rho=.90$); and (4) no differential item functioning with respect to age and gender, but differential item functioning for 4 items in amputees compared with nonamputees.

Conclusions: A hierarchical scale, with excellent scaling characteristics, has been developed for measuring activity limitations in climbing stairs in patients with lower-extremity disorders who live at home. However, measurements should be interpreted with caution when comparisons are made between patients with and without amputation.

Key Words: Activities of daily living; Disability evaluation; Lower extremity; Psychometrics; Questionnaires; Rehabilitation.

© 2004 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

STAIR CLIMBING IS IMPORTANT for maintaining mobility and independence. However, activity limitations in climbing stairs are prevalent, especially in the elderly and in patients with lower-extremity disorders. Despite the importance of stair climbing and the high prevalence of limitations in this activity, little is known about the actual limitations in climbing stairs perceived by patients at home. Furthermore, little is known about the severity of these limitations in subgroups of patients and about the determinants of activity limitations in climbing stairs. Finally, little is known about the severity of the activity limitations indicated by specific statements about these limitations. Most doctors would agree that a statement like "I go up the stairs and always hold onto the banister" indicates a minor limitation, whereas a statement such as "I go up the stairs and am always helped by someone" indicates a severe limitation. But what is implied by statements such as "I go up the stairs but with difficulty" or "I do go up and down stairs but less often"? A prerequisite in addressing these research questions is the availability of a suitable measurement instrument.

We are not aware of any existing measurement instrument that provides a detailed assessment of activity limitations in climbing stairs.¹ Currently available generic measurement instruments²⁻⁷ and disease-specific instruments⁸⁻¹² provide only a global measurement of activity limitations. Although climbing stairs is included in many of these instruments, this activity is often addressed by 1 or 2 items only.

Because existing measurement instruments were not adequate to provide a detailed assessment of activity limitations in climbing stairs, we developed a new scale with the following properties: (1) the scale should be suitable for the assessment of patients with lower-extremity disorders living at home; (2) the scale must provide a detailed measurement of activity limitations in climbing stairs; (3) the scale must be a discriminative index, measuring cross-sectional differences in limitations in climbing stairs between patients or subgroups of patients; and (4) the scale should be hierarchical, which implies that it consists of items indicating different levels of severity in activity limitations.

This article reports on the development and testing of a detailed and hierarchical scale to measure activity limitations in climbing stairs in patients with lower-extremity disorders living at home.

METHODS

Instrument Development

To create a detailed measurement instrument, items were derived from an extensive literature review on climbing stairs.

From the Department of Rehabilitation Medicine (Roorda, Roebroek, Lankhorst) and Institute for Research in Extramural Medicine (Roorda, Lankhorst, Bouter), VU University Medical Center, Amsterdam; Department of Rehabilitation Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam (Roebroek); and Department of Sociology and Social Gerontology, Vrije Universiteit, Amsterdam (van Tilburg), The Netherlands.

Supported by the SGO Health Research Promotion Program.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the author(s) or on any organization with which the author(s) is/are associated.

Reprint requests to Leo D. Roorda, MD, PT, VU University Medical Center, Dept of Rehabilitation Medicine, PO Box 7057, 1007 MB Amsterdam, The Netherlands, e-mail: ld.roorda@vumc.nl.

0003-9993/04/8506-8242\$30.00/0

doi:10.1016/j.apmr.2003.11.018

A first-draft version of the scale was subjected to the opinions of experts (physicians, physical and occupational therapists, sociologists), and pretested. A second draft version was tested in 345 patients with orthopedic and rheumatologic disorders of the lower extremity,^{13,14} resulting in rewording of some items. The final version of the instrument that was tested was a self-administered questionnaire consisting of 15 items, formulated in behavioral terms. The items refer to what a patient actually does, and not to what a patient thinks that he/she can do. Dichotomous response options (YES box marked/YES box not marked) were chosen to facilitate interpretation. A summary of the patient instructions and the final list of items can be found in appendix 1.

Participants

To test the measurement instrument, patients were recruited from several studies of subjects with lower-extremity disorders living at home. First, patients with chronic stroke were sampled from a community-based cohort study,¹⁵ at the 2-year follow-up measurement. Second, patients with poliomyelitis were sampled from a community-based cohort study,¹⁶ also at the 2-year follow-up measurement. Third, outpatients with hip or knee osteoarthritis (OA) undergoing arthroplasty were sampled from 2 general hospitals and 1 university hospital in an ongoing inception cohort study, directly after placement on the waiting list.¹⁷ Fourth, outpatients with lower-extremity amputation undergoing multidisciplinary rehabilitation treatment were sampled from an ongoing inception cohort study, at the end of rehabilitation treatment. In addition, outpatient amputees were sampled from 2 orthopedic workshops. Fifth, patients with lower-extremity complex regional pain syndrome type I (CRPS I) were sampled from the anesthesiology and surgical outpatient clinics of 2 university hospitals.^{18,19} Some of these patients were also participants in a randomized controlled trial,¹⁸ and these patients were sampled at the first measurement. Sixth, outpatients with diabetic foot problems (with and without foot ulcers) from a rehabilitation center were sampled from the cases in an ongoing case-control study.²⁰ In addition, outpatients with diabetic foot problems receiving podiatric care from a general and a university hospital were sampled from an observational study,²¹ at the first measurement. All patients completed the questionnaire and provided additional information about their age and gender.

Analysis

Scalability and invariant item ordering. Scalability of an item set implies that it can be used for the measurement of patients, whereas invariant item ordering implies that the (hierarchical) ordering of the items is the same for all patients. To investigate the scalability and the ordering of the items, Mokken scale analysis²²⁻²⁵ was used in this study. Mokken scale analysis can be viewed as a nonparametric approach to the item response theory (IRT). According to the IRT, the abilities of patients are latent traits. Mokken scale analysis provides only ordinal information about the location of patients and items on the scale of the latent trait. Within the framework of Mokken scale analysis, the scalability of an item set is studied by investigating the fit of the monotone homogeneity (MH) model, whereas the invariant ordering of the items is studied by investigating the fit of the double monotonicity (DM) model.

Evaluation of the MH model. The fit of the MH model is evaluated by calculating the scalability coefficient H .^{23,24} Scalability coefficient H 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 the coefficients of scalability for all item pairs (H_{ij}) are positive, whereas the scalability coefficients for the items in relation to the scale at issue (H_i) and for the scale (H) are at least .30. Higher values for H_i and H imply fewer violations and thus a better scale. A rule of thumb is that a scale is considered to be weak if $.30 \leq H < .40$. Medium scalability is obtained if $.40 \leq H < .50$. A scale is considered to be strong if H is equal to or greater than .50.

Evaluation of the DM model. The evaluation of the fit of the DM model with the data starts with investigating the fit of the MH model and continues with calculating the coefficient H^T .^{23,24} The coefficient H^T is a global indicator of the degree to which the ordering of the items is invariant across the latent trait. Criteria for invariant item ordering are met when the percentage of negative coefficients at the level of the individual patients (H_{ij}^T) is less than 10% and the coefficient for the total set of patients (H^T) is at least .30. The larger the H^T , the greater the confidence that can be assigned to the invariant ordering of items across the latent trait.

Reliability. Reliability concerns the degree of repeatability of the sum score. The intratest reliability was quantified by calculating the reliability coefficient ρ .^{23,24} A reliability coefficient of .90 or more is recommended for decisions about individual patients.²⁶

Robustness. Robustness concerns the scalability of the item set and the invariant item ordering within subgroups of patients. In our study, robustness was investigated with respect to age, gender, and diagnosis. For the purpose of investigating robustness with respect to age, the study population was dichotomized on the basis of the median age (<61.5y vs $\geq 61.5y$). Because good estimates of scalability require large groups of patients (at least 200, but preferably at least 500),²³ robustness with respect to diagnosis was investigated only for the subgroups of patients with OA and amputation and for a rest group consisting of patients with the other diagnoses. Scalability and invariant item ordering within subgroups of patients were evaluated, as described, for all patients.

Differential item functioning. Differential item functioning (DIF) addresses the issue of making valid comparisons between subgroups of patients. First, suppose that 2 subgroups of patients—for instance, OA patients and amputees—have a similar severity of activity limitations. Second, suppose that an item functions differently in OA patients and amputees—for instance, that it indicates less activity limitations in amputees. In such a case, amputees will more often respond positively to this item, and they will tend to have higher scores than the OA patients. So, an item that functions differently in subgroups of patients causes differences in subgroup scores, even when the subgroups have a similar severity of activity limitations. As a consequence, DIF impedes valid comparison between these subgroups. In this study, DIF was investigated for age, gender, and diagnosis. The subgroups were the same as described for the analysis of robustness. DIF is visualized in a scatterplot.

RESULTS

Participants

Of the 822 patients enrolled in our study, 63 stated that they did not climb stairs at all. These 63 patients were excluded from the scale analysis. As a consequence, data on 759 patients were included in the scale analysis. The mean age $\pm$ standard deviation of these patients was 59.8 ± 15.0 years, and the gender of 48% ($n=364$) of the patients was male. Patients had the following diagnoses: stroke ($n=65$)¹⁵; poliomyelitis ($n=89$)¹⁶; hip OA ($n=203$)¹⁷; knee OA ($n=79$); transfemoral amputation

Table 1: Coefficient of Scalability of the 15 Items, and Coefficients of Scalability, Invariant Item Ordering, Intratest Reliability, and Median (IQR) Sum Scores of the Scale of Activity Limitations in Climbing Stairs for All Patients and Subgroups of Patients, Defined by Age, Gender, and Diagnosis

Item*	Age						Gender				Diagnosis					
	All (N=759)		Young (n=380)		Old (n=379)		Male (n=364)		Female (n=395)		OA (n=282)		Amp (n=192)		Other (n=285)	
	Mean [†]	$H_i^{\dagger\dagger}$	Mean	H_i	Mean	H_i	Mean	H_i	Mean	H_i	Mean	H_i	Mean	H_i	Mean	H_i
Minor limitations[§]																
Down: hold on to the banister (10)	.83	.59	.80	.56	.86	.64	.82	.60	.84	.59	.88	.63	.84	.59	.76	.55
Up: hold on to the banister (4)	.82	.58	.79	.52	.84	.64	.81	.55	.83	.60	.87	.60	.82	.57	.76	.54
Up: it takes longer (1)	.66	.62	.67	.67	.66	.57	.67	.64	.66	.60	.74	.56	.63	.62	.61	.64
Down: it takes longer (7)	.61	.58	.61	.61	.62	.55	.61	.57	.62	.59	.70	.53	.61	.61	.53	.58
Down: with difficulty (9)	.56	.53	.56	.58	.56	.47	.56	.52	.57	.53	.68	.47	.45	.50	.52	.58
Up: with difficulty (3)	.56	.52	.58	.56	.54	.49	.55	.53	.57	.52	.68	.48	.39	.46	.55	.60
Up and down: less often (13)	.50	.42	.50	.46	.50	.38	.46	.42	.53	.42	.56	.38	.49	.44	.44	.42
Up: in a different way (2)	.50	.45	.46	.46	.53	.43	.47	.46	.52	.43	.52	.42	.64	.38	.37	.51
Down: in a different way (8)	.43	.44	.39	.44	.47	.43	.41	.45	.45	.43	.45	.43	.56	.42	.32	.44
Up and down: avoid both (14)	.39	.45	.42	.47	.35	.43	.36	.44	.41	.45	.41	.45	.33	.45	.40	.44
Up and down: less stairs/floors (15)	.29	.43	.31	.42	.26	.43	.27	.38	.31	.46	.28	.45	.27	.46	.30	.39
Up: use a walking aid (5)	.14	.45	.11	.55	.18	.39	.15	.34	.14	.56	.15	.35	.20	.39	.11	.61
Down: use a walking aid (11)	.14	.51	.10	.59	.17	.45	.14	.40	.14	.61	.15	.41	.17	.51	.11	.60
Up: am helped by someone (6)	.04	.42	.02	.51	.05	.38	.03	.40	.04	.43	.03	.27	.05	.60	.04	.37
Down: am helped by someone (12)	.03	.39	.02	.33	.05	.40	.03	.43	.04	.35	.03	.21	.05	.62	.03	.31
Severe limitations																
Coefficient of scalability $H^{\parallel}$	.50		.52		.47		.49		.50		.46		.49		.52	
Coefficient of invariant item ordering H^{∇}	.58		.61		.55		.57		.59		.63		.59		.56	
Percentage of negative H_a^{∇} values**	2.9		2.3		3.4		2.7		3.0		2.2		3.4		4.0	
Coefficient of reliability $\rho^{\dagger\dagger}$	.90		.89		.88		.88		.90		.88		.90		.90	
Sum score median ^{††}	7		7		7		7		7		7		7		6	
Sum score IQR	4-9		3-9		4-9		4-9		4-9		5-10		4-9		2-9	

Abbreviations: Amp, amputation; IQR, interquartile range.

* Abbreviated item text. The numbers in parentheses refer to the nonabbreviated items in appendix 1.

[†] Proportion of patients responding positively to the items.[‡] Coefficient of scalability H_i of the items in relation to the scale (range, 0–1 under the MH model). A minimum value of $H_i=.30$ is recommended.[§] Severity of activity limitations increases in the scale from top to bottom.^{||} Coefficient of scalability H for the scale (range, 0–1 under the MH model). $H \geq .50$ indicates a strong scale; $.40 \leq H < .50$ indicates a moderate scale.[¶] Coefficient of invariant item ordering H^{∇} for the total set of patients. A minimum value of $H^{\nabla}=.30$ is recommended.^{**} Percentage of negative coefficients of invariant item ordering H_a^{∇} at the level of the individual patients. A percentage of $<10\%$ is recommended.^{††} Reliability coefficient ρ for the scale (range, 0–1 under the MH model). Coefficient $\rho \geq .80$ is recommended for comparisons between patient groups, and $\rho \geq .90$ is recommended for decisions about individual patients.^{††} Sum scores range from 0–15.

(n=74); knee disarticulation amputation (n=12); transtibial amputation (n=89); other amputation (n=17); CRPS I (n=69)^{18,19}; and diabetic foot problems (n=62).^{20,21}

Mokken Scale Analysis

MH model. The scale criteria were met (table 1), which indicates that the item set can be used for the measurement of patients. The coefficient of scalability for all item pairs (H_{ij}) was positive, the scalability coefficient for the items in relation to the scale at issue (H_i) was .39 or greater, and the scalability coefficient of the scale (H) was .50, indicating a strong scale.

DM model. The criteria for invariant item ordering were met, which indicates that the (hierarchical) ordering of the items, from items indicating minor limitations to items indicating severe limitations, was the same for the whole range of patients with minor to severe limitations. The coefficient for the total set of patients (H^{∇}) was .58, while the percentage of negative coefficients at the level of the individual patients (H_a^{∇}) was 3% (table 1).

Reliability. The intratest reliability coefficient ρ was .90, which indicates that the intratest reliability was good enough for decisions about individual patients.

Robustness. Robustness with respect to age, gender, and diagnosis was satisfactory (table 1), which indicates that the item set can be used for measurement within the subgroups of patients and that the item ordering is invariant within these subgroups. For the subgroup of OA patients only, items 6 and 12 showed misfit of the MH model ($H_i < .30$).

Differential item functioning. No DIF was found for age or gender, and some DIF was found for diagnosis. This indicates that valid comparisons can be made between younger and older patients, and between male and female patients, whereas comparisons between diagnostic groups should be made with some caution. The DIF with respect to diagnosis for some items (2, 3, 8, 9, 13, 14) tended to be mainly attributable to amputees. Consequently, in another analysis, the study population was dichotomized (amputees vs nonamputees). This analysis revealed DIF with respect to the items 2, 3, 8, and 9 (fig 1).

DISCUSSION

The objective of our study was to develop and test a measurement instrument to provide a detailed assessment of activity limitations in climbing stairs for patients with lower-extremity disorders living at home. To test the instrument on patients living at home, only outpatients were sampled for this study. In addition, and in order to test the instrument on patients with different lower-extremity disorders, the patients who were sampled had different disorders of the lower extremity, originating from different parts of the body: the brain (stroke), the peripheral nerve (poliomyelitis, diabetic foot problems), the hip (OA), the upper leg (transfemoral amputation), the knee (OA, knee disarticulation amputation), the lower leg (transtibial amputation), and the foot (CRPS I, diabetic foot problems, foot amputations).

Fig 1. Mean item scores for amputees and nonamputees. Mean item scores for amputees are on the x axis and for nonamputees, on the y axis. Low mean item scores designate the items indicating severe limitations, and high mean item scores designate the items indicating minor limitations. The numbers near the data points refer to the item numbers. For a full explanation of the items, see appendix 1. An identity line is drawn through the origin with a slope of 1. Items with major and significant deviations from the identity line demonstrate DIF. *Items with DIF.

The suitability of the item set for the measurement of patients was confirmed by demonstrating good fit of the MH model for all and within subgroups of patients defined by age, gender, and diagnosis. For the subgroup of OA patients only, 2 items showed misfit of the model ($H_i < .30$), which might be explained by the small number of patients in this subgroup who responded positively to these 2 items (about being helped by someone when climbing stairs). In such a case, violation of the model, caused by an unusual answer from a few patients, can easily result in a low H_i coefficient.

Invariant item ordering, and thus a hierarchy, was confirmed by demonstrating good fit of the DM model for all, and within subgroups of patients. The items can therefore be ordered from items indicating slight limitations to items indicating severe limitations. As expected, the items about holding onto the banister indicate minor limitations, whereas the items about being helped by someone indicate severe limitations. Generally speaking, there were no great differences between the limitations indicated by the items about going upstairs and the items about going downstairs, such as the items 1 and 7. Because of the invariant item ordering, a patient with a sum score of 4 will generally respond positively to the first 4 items indicating minor limitations (items 10, 4, 1, 7) and negatively to the other 11 items indicating more severe limitations.

There are some concerns about the validity of comparisons between amputees and nonamputees, because 4 items demonstrated DIF. Items 2 and 8 indicate less severe activity limitations for amputees than for nonamputees. The finding that these items about climbing stairs in a different way (eg, pulling up or down 1 leg at a time) indicated less activity limitations in amputees may be explained by the use of a prosthesis that forces amputees to climb stairs in a different way, even if they

have only minor activity limitations. Items 3 and 9 indicate more severe activity limitations in climbing stairs for amputees than for nonamputees. This finding might be explained by the fact that the nonamputee group mainly consisted of patients with painful disorders (OA, CRPS I, diabetic foot problems). Complaints about pain and difficulty are often related.²⁷ Patients with a painful disorder may immediately start reporting "difficulty" when they experience pain, which might be in a phase in which they have only minor activity limitations. Although items with DIF do not always produce poor measurements,²⁸ the use of only the 11 items without DIF is recommended when comparisons are made between amputees and nonamputees.

In this study, test-retest reliability and construct validity were not investigated. Perez et al¹⁹ have demonstrated a satisfactory test-retest reliability (intraclass correlation coefficient = .87) of this scale in CRPS I patients (N=21). In addition, Meijer et al²⁰ provided some evidence for the construct validity of the scale by demonstrating that diabetic patients (N=38) with a high risk of developing foot complications had more severe activity limitations in climbing stairs. Test-retest reliability and construct validity will, however, have to be investigated more thoroughly.

CONCLUSIONS

A measurement instrument with excellent scaling characteristics has been developed to provide a detailed assessment of activity limitations in climbing stairs in patients with different lower-extremity disorders living at home. However, the results of the measurements made with this instrument should be interpreted with some caution when comparing amputees and nonamputees.

Acknowledgments: We thank Dr. Ivo W. Molenaar for his comments on the manuscript. Other members of the Measuring Mobility Study Group are Fred A. de Laat, MD, Roessingh Research and Development, Enschede, and Rehabilitation Center Tolbrug, 's Hertogenbosch; Ide C. Heyligers, MD, PhD, VU University Medical Center, Amsterdam; Jacob J. Caron, MD, Onze Lieve Vrouwe Gasthuis, Amsterdam; Bart J. Burger, MD, PhD, Medisch Centrum Alkmaar, Alkmaar; Frans Nollet, MD, PhD, VU University Medical Center, Amsterdam; Irene E. Stover-van Herk, MSc, University of Groningen, Groningen; Roberto S. Perez, PhD, VU University Medical Center, Amsterdam; Anne M. Boonstra, MD, PhD, Rehabilitation Center Revalidatie Friesland, Beetsterzwaag; Jan W. Meijer MD, PhD, Rehabilitation Center Tolbrug, 's Hertogenbosch, University of Groningen, Groningen; P. Mieke Rijken, Netherlands Institute of Health Services Research, Utrecht, The Netherlands.

APPENDIX 1

Please answer YES to every statement that **both** applies to your current situation and is connected with your health.

Going Up and Down the Stairs

	YES
1. I go up the stairs but it takes longer.	☐
2. I go up the stairs but in a different way, eg, I pull up 1 leg at a time.	☐
3. I go up the stairs but with (some) difficulty.	☐
4. I go up the stairs and (almost) always hold onto the banister.	☐
5. I go up the stairs and (almost) always use a walking aid, eg, a walking stick or a crutch.	☐
6. I go up the stairs and am (almost) always helped by someone.	☐

7. I go down the stairs but it takes longer. ☐
8. I go down the stairs but in a different way, eg, I pull down 1 leg at a time or I go down "backward." ☐
9. I go down the stairs but with (some) difficulty. ☐
10. I go down the stairs and (almost) always hold onto the banister. ☐
11. I go down the stairs and (almost) always use a walking aid, eg, a walking stick or a crutch. ☐
12. I go down the stairs and am (almost) always helped by someone. ☐
13. I do go up and down the stairs but less often. ☐
14. I do go up and down the stairs but I avoid them. ☐
15. I do go up and down the stairs but less stairs/floors. ☐

References

1. Van Iersel MB, Olde Rikkert MG, Mulley GP. Is stair negotiation measured appropriately in functional assessment scales? *Clin Rehabil* 2003;17:325-33.
2. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. *Md State Med J* 1965;14:61-5.
3. Hunt SM, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281-6.
4. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
5. De Bruin AF, Diederiks JP, de Witte LP, Stevens FC, Philipsen H. The development of a short generic version of the Sickness Impact Profile. *J Clin Epidemiol* 1994;47:407-18.
6. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987;1:6-18.
7. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
8. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969;51:737-55.
9. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982;9:789-93.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
11. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop* 1989;Nov(248):13-4.
12. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis Rheum* 1992;35:1-10.
13. Roorda LD, Roebroek ME, Lankhorst GJ, van Tilburg T, Bouter LM. Measuring functional limitations in rising and sitting down: development of a questionnaire. *Arch Phys Med Rehabil* 1996;77:663-9.
14. Roorda LD, Roebroek ME, Lankhorst GJ, van Tilburg TG. De vragenlijst loopvaardigheid: hierarchische schalen om beperkingen in het opstaan en lopen te meten. *Revalidata* 1996;18:34-8.
15. Van Herk IE, Arendzen JH, Rispens P. Ten-metre walk, with or without a turn? *Clin Rehabil* 1998;12:30-5.
16. Nollet F, Beelen A, Prins MH, et al. Disability and functional assessment in former polio patients with and without postpolio syndrome. *Arch Phys Med Rehabil* 1999;80:136-43.
17. Roorda LD, Jones CA, Waltz M, et al. Satisfactory crosscultural equivalence of Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004;63:36-42.
18. Perez RS, Zuurmond WW, Bezemer PD, et al. The treatment of complex regional pain syndrome type I with free radical scavengers: a randomized controlled study. *Pain* 2003;102:297-307.
19. Perez RS, Roorda LD, Zuurmond WW, Bannink II, Vranken JH, de Lange JJ. Measuring perceived activity limitations in lower extremity Complex Regional Pain Syndrome type I (CRPS I): test-retest reliability of two questionnaires. *Clin Rehabil* 2002;16:454-60.
20. Meijer JW, Trip J, Jaegers SM, et al. Quality of life in patients with diabetic foot ulcers. *Disabil Rehabil* 2001;23:336-40.
21. Rijken PM, Dekker J, Lankhorst GJ, et al. Podiatric care for diabetic patients with foot problems: an observational study. *Int J Rehabil Res* 1999;22:181-8.
22. Mokken RJ. Nonparametric models for dichotomous responses. In: Van Der Linden WJ, Hambleton RK, editors. *Handbook of modern item response theory*. New York: Springer; 1997. p 351-67.
23. Molenaar IW, Sijtsma K. User's manual MSP5 for Windows. Groningen: iecProGAMMA; 2000.
24. Sijtsma K, Molenaar IW. Introduction to nonparametric item response theory. Thousand Oaks: Sage; 2002.
25. Molenaar IW. Parametric and nonparametric item response theory models in health related quality of life measurements. In: Mesbah M, Cole BF, Lee ML, editors. *Statistical methods for quality of life studies: design, measurements and analysis*. Dordrecht: Kluwer Academic; 2002. p 143-54.
26. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
27. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care Res* 1999;12:331-5.
28. Roznowski M, Reith J. Examining the measurement quality of tests containing differentially functioning items: do biased items result in poor measurement? *Educ Psychol Meas* 1999;29:248-69.