VU Research Portal

Responsiveness of the Rehabilitation Activities Profile and the Barthel Index

van Bennekom, C.A.M.; Jelles, F.; Lankhorst, G.J.; Bouter, L.M.

published in

Journal of Clinical Epidemiology
1996

DOI (link to publisher)

[10.1016/0895-4356\(95\)00559-5](https://doi.org/10.1016/0895-4356(95)00559-5)

document version

Publisher's PDF, also known as Version of record

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

citation for published version (APA)

van Bennekom, C. A. M., Jelles, F., Lankhorst, G. J., & Bouter, L. M. (1996). Responsiveness of the Rehabilitation Activities Profile and the Barthel Index. *Journal of Clinical Epidemiology*, 49(1), 39-44. [https://doi.org/10.1016/0895-4356\(95\)00559-5](https://doi.org/10.1016/0895-4356(95)00559-5)

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

Responsiveness of the Rehabilitation Activities Profile and the Barthel Index

Coen A. M. van Bennekom,^{1*} Frank Jelles,¹ Gustaaf J. Lankhorst,¹ and Lex M. Bouter²

¹FREE UNIVERSITY HOSPITAL, DEPARTMENT OF REHABILITATION MEDICINE, AMSTERDAM, THE NETHERLANDS, AND ²VRIJE UNIVERSITEIT, FACULTY OF MEDICINE, DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS, AMSTERDAM, THE NETHERLANDS

ABSTRACT. The goal of this study was to compare the responsiveness for clinically meaningful change over time of a newly designed functional status scale, the Rehabilitation Activities Profile (RAP), with the more frequently used Barthel Index (BI). Four techniques for the quantification of responsiveness were utilized: effect sizes, *p*-values, *t*-statistics and ROC curves. The patient's return home was chosen as external criterion. An inception cohort of stroke patients was followed during 26 weeks. All patients still hospitalized on the 14th day after stroke were included. The functional assessments took place at 2, 3, 4, 8, 12, and 26 weeks after stroke. The patients were visited at the hospital, home, nursing home, or rehabilitation center. Of the 125 patients included in the study, 18 patients died during the observation period, 2 patients were lost to follow-up, and 1 patient refused to cooperate after 12 weeks. After 26 weeks, 104 patients remained for analysis. Three time periods were discerned: 2 to 12 weeks (early response), 12 to 26 weeks (late response), and 2 to 26 weeks after stroke (overall response). The effect sizes of the RAP were consistently higher on all three time periods than those of the BI. The *p*-value of the overall response mean change score of the RAP appeared to discriminate between patients returning home and those not returning home, whereas the BI failed on this point (*p* = 0.004 vs. 0.496). Using *t*-statistics, the RAP showed a higher efficiency in expressing change on all time periods (relative efficiency = 1.42, 1.77, and 1.43, respectively). The receiver operating characteristic surface area of the RAP score was higher than the area of the BI score (0.74 and 0.59, respectively for the early response period). In conclusion, all results seem to indicate that the RAP is more responsive than the BI when returning home is chosen as an external criterion. J CLIN EPIDEMIOL 49:1:39–44, 1996.

KEY WORDS. Functional assessment, rehabilitation, responsiveness

INTRODUCTION

The goal of most treatments in rehabilitation medicine is to induce an improvement of the functional status of the patient. Many assessment methods are developed to measure functional status [1]. The need to establish their reliability and validity is generally accepted. However, for assessment methods that are supposed to measure within-person change over time, these cross-sectional properties are not sufficient. An essential clinimetric property of evaluative methods is the ability to measure clinically meaningful change over time, called sensitivity to change or responsiveness [2]. The term responsiveness is preferable to sensitivity to change, because sensitivity is also used to describe the performance of a diagnostic test in the medical and epidemiologic literature. Responsiveness can be defined as the ability to detect minimal clinically important differences [3,4]. An instrument showing a relatively small change in score while the patient improves dramatically is poorly responsive. On the other hand, if the instrument shows some change while the improvement occurring is quite small, the method is called highly responsive.

There is no consensus on the appropriate strategy to quantify responsiveness. Most techniques use the magnitude of a statistical test (e.g., *t*-statistics, *p*-values) or the ratio of the difference between groups to

the variability within groups. These techniques share the requirement for an external criterion to define the improved patient group. This external criterion should actually define the minimal clinically important difference. Another method to ensure that the patients in question improve is by applying a treatment of known efficacy [2].

This study was designed to compare the responsiveness of a newly designed functional status measurement with a frequently used instrument: the Rehabilitation Activities Profile (RAP) [5] versus the Barthel Index (BI) [6].

METHODS

Patient Population

Patients in this study were those enrolled in a prospective, descriptive study on the prognosis of stroke. Consecutive stroke patients of four departments of neurology were involved: Free University Hospital (September 1992 to September 1993), Sint Lucas Hospital Amsterdam (January 1993 to September 1993), University Hospital Groningen (March 1993 to June 1993), and Martini Hospital Groningen (March 1993 to June 1993). Only those patients who were still hospitalized on the 14th day after stroke were included. This two-week criterion was used so as to include a medically stable group with a well-established diagnosis. Patients suffering from a subarachnoid hemorrhage were excluded. After informed consent was obtained, clinical, sociodemographic, and functional health status data were collected at 2, 3, 4, 8, 12, and 26 weeks after stroke.

*All correspondence should be addressed to: Dr. Coen AM van Bennekom, Free University Hospital, Department of Rehabilitation Medicine, PO Box 7057, 1007 MB Amsterdam, The Netherlands.

(Received in revised form 21 March 1995)

Functional Status Measures

REHABILITATION ACTIVITIES PROFILE. The RAP* was constructed by our department as an assessment method that facilitates (1) the selection of patients for rehabilitation, (2) the monitoring of progress, and (3) the expression of a functional prognosis. The method is described elsewhere in detail [5]. The reliability of the method has been tested extensively and has proven to be good [7]. To condense information, sum scores are constructed. Evidence has been found for its use as a predictive, discriminative, and evaluative index.

BARTHEL INDEX. The BI is a frequently used measure of mobility and personal care [6]. It consists of the assessment of ten activities that relate to the patient's dependency of help. The BI uses a sum score. Its reliability and validity has been tested extensively [8–11].

Appendix 1 shows the concordance between the BI and the RAP domains of mobility and personal care, and the construction of the sum scores of both methods.

Data Analysis

The sum score of the RAP domains of mobility and personal care ranges from 0 to 30, while that of the BI ranges from 0 to 20. The lower the RAP sum score, the less disabled the patient, while the opposite is true for the BI. For the analysis we reversed the RAP sum score (by calculating 30 minus the sum score of the RAP domains of mobility and personal care) and standardized all scores to a scale of 0 to 100.

To conform with other stroke studies, we chose to analyse three time points: 2 weeks, 3 months, and 6 months after stroke. Means and standard deviations were calculated for these time points. Differences between serial assessments were calculated by subtracting the later score from the earlier score. Thus, a positive score difference indicates improvement while a negative difference indicates deterioration. Three response periods were calculated: (1) 12-week score minus 2-week score (early response), (2) 26-week score minus 12-week score (late response), and (3) 26-week score minus 2-week score (overall response).

Methods for Assessing Responsiveness

Four different approaches to quantify responsiveness were utilized (Table 1).

EFFECT SIZE. Effect size relates the magnitude of the change score to the variability in score [12]. In general, the effect size is calculated by taking the mean change found in a particular variable and dividing it by the standard deviation of that variable. However, there is some controversy regarding which standard deviation to take. Cohen [12] used the standard deviation of the mean change score of the same group. Liang *et al.* [13] did the same but used the term standardized response mean to avoid the ambiguity in the term effect size. Kazis *et al.* [14] used the standard deviation of the baseline score. Guyatt *et al.* [4,15] advocate the standard deviation of the mean change score in stable patients. They relate clinically meaningful change scores in a functional status measurement to the variability of scores in stable subjects.

In this study, all three methods of effect size calculation are calculated. In general, if two assessment methods that are supposed to

measure the same response phenomenon have equal effect size values, they are statistically equivalent for that response. A higher effect size indicates greater responsiveness. The interpretation of the magnitude of the effect size is often based on Cohen's rule-of-thumb: an effect size of 0.2 represents a small effect of an intervention, 0.5 represents a moderate effect, and 0.8 or greater represents a large effect [12].

p-VALUES. Guyatt *et al.* [16] used the *p*-value as an index of responsiveness. A responsive assessment method should be able to differentiate between two different groups: larger differences in score are expected in the group that improved on an external criterion than in the group without improvement. The *p*-value is calculated using an unpaired *t*-test. The smaller the *p*-value of the difference between the mean scores, the greater the responsiveness.

RELATIVE EFFICIENCY. Liang *et al.* [17] compared pre- and posttreatment scores for a treatment of known efficacy as indicators of responsiveness. The resulting *t*-statistic can be used to compare competing assessment methods. The method with the largest paired *t*-statistic is judged to be the most responsive. The relative efficiency is calculated by dividing the *t*-statistics of two assessment methods.

RECEIVER OPERATOR CHARACTERISTIC CURVES. Deyo and Centor [3] view functional status scales as diagnostic tests for discriminating between improved and unimproved patients. The receiver operating characteristic (ROC) curve describes a scale's ability to detect improvement (or failure to improve) using an external criterion. The ROC consists of a graph of true positive (sensitivity) versus false positive (1-specificity) for each of several cut-off points in change score. The area under a ROC curve can be interpreted as the probability of correctly identifying the improved patient from randomly selected pairs of improved and unimproved patients [18]. This area theoretically ranges from 0.5 (no accuracy in discriminating improved from unimproved) to 1.0 (perfect accuracy).

All methods need an external criterion to distinguish between the improved and the nonimproved group. We have chosen living arrangement as external criterion. The domains of mobility and personal care, as measured by the RAP and the BI, represent the patient's functional independence. One would expect that patients who live at home 26 weeks after stroke would show a greater change in score than patients who are unable to go home. Thus, the chosen external criterion in this study is whether the patients are living at home 26 weeks after stroke. The amount of change in the RAP and the BI between different measurement points is compared by applying the four approaches to responsiveness quantification discussed above.

RESULTS

The study population characteristics are shown in Table 2. Of the 125 patients included in the study, 18 patients died during the observation period, 2 patients were lost to follow-up, and 1 patient refused to cooperate after 12 weeks. Of the remaining 104 patients, 57 (55%) were living at home 26 weeks after their stroke; 47 patients were not living at home, of whom 41 (39%) lived in a nursing home, 3 (3%) were still in the rehabilitation center, and 3 (3%) were being cared for by their family. Figure 1 shows the living arrangements during the observation period.

Table 3 shows the mean sum scores for all time points, and Table 4 shows the change scores for the RAP and the BI at the three discerned response periods. The greater part of improvement took place in the first 12 weeks (the so-called early response). The differences between the later score and the earlier score were statistically significant ($p \leq$

*A copy of the RAP and its manual is available on request from the first author.

TABLE 1. Different approaches to the quantification of responsiveness

Name	Author(s)*	Method
Effect size	Cohen [12]	Mean change score/SD of mean change score
	Liang <i>et al.</i> [13]	
	Katz <i>et al.</i> [21]	
	Kazis <i>et al.</i> [14]	Mean change score/SD of baseline score
	Fitzpatrick <i>et al.</i> [22]	
p-values	Guyatt <i>et al.</i> [4,15]	Mean change score/SD of mean change score in stable patients
	Guyatt <i>et al.</i> [16]	Unpaired <i>t</i> -test on differences between mean scores of improved and non-improved patients
Relative efficiency	Liang <i>et al.</i> [17]	Ratio of the paired <i>t</i> -statistic on difference score of improved patients of two competing instruments
	Deyo and Centor [3]	
ROC curve	Deyo and Centor [3]	Sensitivity and specificity of change in score as related to the dichotomized outcome
	MacKenzie <i>et al.</i> [23]	

Abbreviations: ROC = receiver operator characteristic; SD = standard deviation.

*Includes studies in which the index of responsiveness at issue is applied.

TABLE 2. Study population (N = 125)

Details of patients	
Mean age (SD)	73.1 (10.7)
Gender	
female	58 (46%)
male	67 (54%)
Living with partner	
no	54 (43%)
yes	71 (57%)
Details of stroke	
History	
first stroke	101 (81%)
recurrence	24 (19%)
Localization	
right hemisphere	54 (43%)
left hemisphere	57 (46%)
brainstem	11 (9%)
cerebellar	3 (2%)
Type	
infarction	104 (83%)
hemorrhage	21 (17%)

Abbreviation: SD = standard deviation.

0.01) for all response periods. There is a striking difference in the late response change score between patients at home and those not at home 26 weeks after stroke. Patients not at home improve in functional status considerably more than patients at home. The difference in baseline between the two groups is obvious, the mean 26-week score of patients not at home does not even equal the baseline of patients at home 26 weeks after stroke.

A drawback of the BI is the occurrence of a ceiling effect: the maximal score has already been reached before functional improvement has stopped [19]. This effect is also observed in this study. Twelve weeks after stroke the mean BI score is 93.8 for the group patients at home after 26 weeks (Table 4). This explains the small standard deviations of the BI at 12 and 26 weeks. There is not much room left for further improvement. The RAP scores have used only three-quarters of its scale length at 12 weeks after stroke.

Effect Size

The alternative calculations of effect size are presented in Table 5. As expected, the effect sizes of the early response are much higher than those of the late response. Independent of the calculation method, the effect size of the RAP is consistently higher than that of the BI.

FIGURE 1. Living arrangements 2, 4, 8, 12, and 26 weeks after stroke.

TABLE 3. Mean sum scores (SD) of the Rehabilitation Activities Profile (RAP) domains of mobility and personal care and the Barthel Index (BI) for all time points

	RAP					
	2-week score	3-week score	4-week score	8-week score	12-week score	26-week score
At home (N = 57)	43.2 (20.8)	52.1 (21.7)	57.4 (22.3)	69.2 (17.4)	74.1 (15.7)	78.4 (15.7)
Not at home (N = 47)	13.3 (13.8)	15.5 (16.2)	18.9 (17.3)	24.5 (19.4)	29.6 (22.1)	37.1 (23.6)

	BI					
	2-week score	3-week score	4-week score	8-week score	12-week score	26-week score
At home (N = 57)	61.8 (29.0)	73.3 (26.9)	78.9 (23.5)	90.2 (12.0)	93.8 (9.1)	96.2 (6.1)
Not at home (N = 47)	19.0 (21.0)	22.0 (23.6)	24.4 (25.2)	33.3 (27.1)	42.7 (30.2)	49.7 (32.4)

Note: All scores have been standardized to a scale of 0 to 100.

Abbreviation: SD = standard deviation.

TABLE 4. Mean change scores (SD) of the Rehabilitation Activities Profile (RAP) domains of mobility and personal care and the Barthel Index (BI) for early, late, and overall responses*

	RAP			BI		
	Early response	Late response	Overall response	Early response	Late response	Overall response
At home (N = 57)	30.9 (17.5)	4.3 (7.5)	35.2 (20.5)	32.0 (25.7)	2.5 (7.6)	34.5 (28.7)
Not at home (N = 47)	16.2 (15.8)	7.5 (10.9)	23.8 (19.2)	23.6 (23.6)	7.0 (14.2)	30.6 (28.3)

Note: All scores have been standardized to a scale of 0 to 100.

*Early response = 12 weeks - 2 weeks; late response = 26 weeks - 12 weeks; overall response = 26 weeks - 2 weeks.

Abbreviation: SD = standard deviation.

TABLE 5. Effect sizes of the Rehabilitation Activities Profile (RAP) domains of mobility and personal care and the Barthel Index (BI) in the group living at home after 26 weeks (N = 57)

	RAP			BI		
	Early response*	Late response*	Overall response*	Early response	Late response	Overall response
cf. Cohen/Liang	1.77	0.57	1.72	1.24	0.33	1.20
cf. Kazis	1.49	0.28	1.69	1.10	0.27	1.19
cf. Guyatt	1.96	0.39	1.83	1.36	0.18	1.22

*Early response = 12 weeks - 2 weeks; late response = 26 weeks - 12 weeks; overall response = 26 weeks - 2 weeks.

TABLE 6. Differences and corresponding *p*-values between the mean change scores in the group living at home and the group not living at home 26 weeks after stroke

	Early response*		Late response*		Overall response*	
	Difference	<i>p</i> -value	Difference	<i>p</i> -value	Difference	<i>p</i> -value
RAP	14.64	0.000	-3.25	0.086	11.39	0.004
BI	8.40	0.086	-4.57	0.052	3.84	0.496

*Early response = 12 weeks - 2 weeks; late response = 26 weeks - 12 weeks; overall response = 26 weeks - 2 weeks.

Abbreviations: RAP = Rehabilitation Activities Profile; BI = Barthel Index.

TABLE 7. Relative efficiency (RE) of the Rehabilitation Activities Profile (RAP) domains of mobility and personal care and the Barthel Index (BI) among patients living at home 26 weeks after stroke

	Early response*		Late response*		Overall response*	
	Paired <i>t</i> -statistic	RE	Paired <i>t</i> -statistic	RE	Paired <i>t</i> -statistic	RE
RAP	13.33	1.42	4.30	1.77	12.95	1.43
BI	9.41	1.00	2.43	1.00	9.08	1.00

Note: All figures of the paired *t*-statistic correspond with $p < 0.001$.

*Early response = 12 weeks - 2 weeks; late response = 26 weeks - 12 weeks; overall response = 26 weeks - 2 weeks.

This indicates that the RAP is more responsive than the BI with living at home 26 weeks after stroke as an external criterion.

p-Values

The differences in mean change score between patients living home and those not living home 26 weeks after stroke and their corresponding *p*-values are given in Table 6. It appears that the overall response change in BI score between the two groups does not differ statistically.

Relative Efficiency

The relative efficiency also indicates that the RAP is more responsive than the BI (Table 7).

ROC Curves

Figure 2 shows the ROC curves of the RAP and the BI change score in the overall response period. The RAP curve is situated above the diagonal, indicating some degree of responsiveness. The BI curve oscillates around the diagonal and seems to have limited responsiveness. For the early response the respective figures were 0.74 and 0.59 for the RAP and BI and for the late response period 0.58 and 0.60.

DISCUSSION

An important feature of evaluative instruments is the degree of responsiveness. Using effect sizes or *t*-statistics to quantify responsiveness, high degrees of responsiveness were found for both the RAP and the BI. The RAP turned out to be the more responsive of the two instru-

ments. When *p*-values were applied, a low responsiveness for the BI was found and a high responsiveness for the RAP. The ROC curve areas did not differ much between the two instruments, except for the early response period.

When comparing the contents of the RAP and the BI, one would indeed expect differences in responsiveness in the direction we found. Originally, the BI was constructed to estimate the amount of nursing care needed in the clinic [6]. This explains why BI only assesses indoor mobility and why incontinence is weighted so strongly (30% of the total score). The RAP assesses walking and using a wheelchair indoors and outdoors, and incontinence has a less important influence on the score. Moreover, the RAP uses a four-point severity scale, which takes into account the difficulty in performing an activity. The fact that the RAP measures outdoor activities as well might explain the higher degree of responsiveness and the absence of a ceiling effect.

In this study, returning home was chosen as the external criterion. In our view this is a tenable choice, because it is the ultimate goal of rehabilitation and is strongly dependent on the functional status. However, it is clearly not the minimal clinically important difference. There are no standardized methods for defining such a minimal difference, which may conceivably differ in different settings [20,21]. Thus our choice of external criterion leads to a certain (but unknown) overestimation of responsiveness. The effect sizes (Table 5) would appear to be quite high, especially taking Cohen's rule-of-thumb into account [12]. Still, our chosen external criterion is an objective one that makes comparison between measures possible.

Another major methodologic problem is the predictive relationship between the initial functional status and the chosen external criterion. In our study, patients with a high functional status score at the baseline returned home despite the fact that their gain in functional status might be small. Evidence for this is found in Table 3, which shows a three-fold difference in baseline score for both instruments between patients returning home and those not returning home. This indicates a clear predictive relationship, which might influence the magnitude of the responsiveness coefficients. So far no statistical technique has been proposed to correct for such baseline differences in the analysis of responsiveness.

There seems to be an imbalance between the ROC curve areas and the other responsiveness measures. Although the chosen criterion showed a rather large clinical difference, the ROC curve areas are not very impressive. This might indicate that this method leads to an underestimation of the responsiveness.

Our study clearly demonstrates that responsiveness calculations should be considered with some caution, especially when respon-

FIGURE 2. ROC curves of the RAP and the BI change scores in the overall response period.

siveness is studied exclusively in the improved group. Changes in the BI score seem not to discriminate between the group of patients who returned home and those who did not (Table 6). This is due to the fact that both outcome groups gain almost the same amount on functional status, although starting from a different baseline score. The RAP also detects functional status gain in both groups, but patients who return home improve more than those not going home, rendering discrimination possible.

CONCLUSION

The concept of responsiveness appears to be difficult to measure, as differences in responsiveness between the RAP and the BI depend to a large extent on the method used. Furthermore, defining the minimal clinically important difference and correcting for baseline differences remain challenges for future research. Despite this, all methods applied seem to support the high degree of responsiveness of the RAP compared with the BI.

We are grateful to J. H. Arendzen, MD, PhD (University Hospital Groningen), C. Bouwsma, MD (Martini Hospital Groningen), G. J. Hazenberg, MD (Free University Hospital Amsterdam), and J. A. L. Vanneste, MD, PhD (Sint Lucas Hospital Amsterdam) for their cooperation. We would especially like to thank Ms. J. Goorhuis, MSc, for her assistance in the northern part of The Netherlands.

This study was supported by a grant from the Dutch Ministry of Welfare, Public Health, and Cultural Affairs (Project I 90.022).

REFERENCES

- Wade DT. Measurement in Neurological Rehabilitation. New York: Oxford University Press; 1992.
- Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chron Dis* 1985; 38: 27-36.
- Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chron Dis* 1986; 39:897-906.
- Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987; 40: 171-178.
- Bennekom CAM van, Jelles F, Lankhorst GJ. Rehabilitation Activities Profile, the ICIDH as a framework for a problem-oriented assessment method in rehabilitation medicine. *Disabil Rehabil* 1995; 17: 169-175.
- Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *MD State Med J* 1965; 14: 61-65.
- Jelles F, Bennekom CAM van, Lankhorst GJ, Sibbel TCJP, Bouter LM. Inter- and intra-rater agreement of the Rehabilitation Activities Profile. *J Clin Epidemiol* 1995; 48: 407-416.
- Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation: analysis of repeated Barthel Index measures. *Arch Phys Med Rehabil* 1979; 60:14-17.
- Wade DT, Collin C. The Barthel Index: a standard measure of physical disability? *Int Disabil Stud* 1988; 10: 64-67.
- Collin C, Wade DT, Davies S, Horne V. The Barthel Index: a reliability study. *Int Disabil Stud* 1988; 10: 61-63.
- Chino N. Efficacy of Barthel Index in evaluating activities of daily living in Japan, the United States, and United Kingdom. *Stroke* 1990; 21(suppl II): 64-65.
- Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press; 1977; 1-27.
- Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990; 28:632-642.
- Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: S178-S189.
- Guyatt GH, Eagle DJ, Sackett B, Willan A, Griffith L, McIlroy W, Patterson CJ, Turpie I. Measuring quality of life in the frail elderly. *J Clin Epidemiol* 1993; 46: 1433-1444.
- Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; 42: 403-408.
- Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985; 28: 542-547.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29-36.
- Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989; 42: 703-709.
- Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures: statistics and strategies for evaluation. *Control Clin Trials* 1991; 12: 142S-158S.
- Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; 30: 917-925.
- Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Importance of sensitivity to change as a criterion for selecting health status measures. *Quality in Health Care* 1992; 1: 89-93.
- MacKenzie CR, Charlson ME, DiGioia D, Kelley K. Can the Sickness Impact Profile measure change? An example of scale assessment. *J Chron Dis* 1986; 39: 429-438.

APPENDIX 1. Sum score construction of the Rehabilitation Activities Profile (RAP) domains of mobility and personal care and its concordance with the Barthel Index (BI)

RAP		BI	
Activity	Score	Activity	Score
Mobility			
Maintaining posture	0-3	-	-
Changing posture	0-3	Transfer	0-3
Walking/using wheelchair*	0-3	Walking	0-3
Climbing stairs	0-3	Stairs	0-2
Using transportation	0-3	-	-
Personal care**			
Eating/drinking	0-3	Feeding	0-2
Washing/grooming	0-3	Grooming	0-1
		Bathing	0-1
Dressing	0-3	Dressing	0-2
Undressing	0-3		
Maintaining continence	0-3	Toilet use	0-2
		Bladder control	0-2
		Bowel control	0-2
total	0-30	total	0-20

*The lowest score of the two items is used in the summation.

**The item 'sleeping' is left out because of low, non-significant correlations with other items (unpublished observations).