

2003

Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials

Alexander W. Dromerick
Washington University School of Medicine in St. Louis

Dorothy F. Edwards
Washington University School of Medicine in St. Louis

Michael N. Diringier
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation

Dromerick, Alexander W.; Edwards, Dorothy F.; and Diringier, Michael N., "Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials." *Journal of Rehabilitation Research and Development*. 40,1. 1-8. (2003).
https://digitalcommons.wustl.edu/open_access_pubs/3708

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.

Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials

Alexander W. Dromerick, MD; Dorothy F. Edwards, PhD; Michael N. Diringer, MD

Department of Neurology and Program in Occupational Therapy Washington University School of Medicine, St. Louis, MO

Abstract—Although most current stroke intervention trials use disability scales to determine outcome, little is known about the sensitivity to change of these scales. The use of a more sensitive measure would increase the statistical power of rehabilitation treatment trials. We applied four well-known disability scales to a group of stroke rehabilitation inpatients to compare sensitivity to change. Ninety-five consecutive admissions to a stroke rehabilitation service were assessed for disability on admission and discharge. Two global scales, the Modified Rankin Scale (MRS) and the International Stroke Trial Measure (ISTM), were compared with two activities of daily living (ADL) scales, the Barthel Index (BI) and the Functional Independence Measure (FIM). We determined the number of patients that each scale detected a clinically significant change in disability. Standardized response means (SRM) and receiver operating characteristic (ROC) analyses were performed. The MRS detected change in 55 subjects, including all who changed on the ISTM; the ISTM detected change in only 23 subjects. The BI detected change in 71 subjects but demonstrated ceiling effects with 26% of subjects scoring >95 . The FIM was most sensitive, detecting change in 91 subjects; no patient achieved a maximum score. The SRM of the FIM was superior to that of the BI (2.18 versus 1.72), and ROC analysis revealed C-statistics of 0.82 for the BI, 0.59 for the MRS, and 0.51 for the ISTM. Global scales were much less sensitive to changes in disability than were ADL scales. Though ADL scales may take longer to administer, their increased sensitivity may make them more useful in treatment trials by allowing fewer subjects to be enrolled.

Key words: cerebrovascular disorders, clinical trials, disability evaluation, outcome assessment (health care), rehabilitation.

INTRODUCTION

The advent of randomized controlled trials in rehabilitation has forced a critical examination of appropriate end points. The recent success of the tissue plasminogen activator (t-PA) trials in acute stroke has led to the expectation that rehabilitation trials will use the same trial end points. Most current trials of early stroke interventions use disability scales to assess outcome and determine efficacy. Many scales are being used; most ongoing trials use either the Modified Rankin Scale (MRS) or the Barthel Index (BI) or both [1,2]. A newer measure, specifically designed for intervention trials, was proposed for use in the International Stroke Trial Measure (ISTM) that examined the efficacy of aspirin or heparin administration [3]. Stroke investigators have chosen these measures because of perceived high clinical utility, high interrater reliability, and ease of use [4–11]. A fourth disability scale, the Functional Independence Measure (FIM), is

Abbreviations: ADL = activities of daily living, BI = Barthel Index, FIM = Functional Independence Measure, ICD = International Classifications of Diseases, ISTM = International Stroke Trial Measure, MRS = Modified Rankin Scale, NINDS = National Institutes of Neurological Disorders and Stroke, ROC = receiver operating characteristic, SRM = standardized response means, t-PA = tissue plasminogen activator, UK-TIA = United Kingdom transient ischemic attack.

Address all correspondence and requests for reprints to Alexander W. Dromerick, MD; Department of Neurology, Campus Box 8111, 660 South Euclid Street, St. Louis, MO 63110; 314-454-7756; fax: 314-454-7759; email: dromericka@neuro.wustl.edu.

widely used in rehabilitation centers and has properties useful for stroke investigators [12].

While much is known about the validity and interrater reliability of these four scales, less is known about their capability to detect change in degree of disability. Since the statistical power of a stroke trial is determined in part by the responsiveness to change of the outcome measure employed, a more sensitive measure allows efficacy to be tested with a smaller number of patients [13,14]. Some have predicted that stroke patients will someday be treated with a combination of different treatments [15,16], each with a variable but measurable effect. If so, highly sensitive scales will be needed to assess treatment efficacy. Thus, a better understanding of the relative sensitivities of specific scales might lead to improvements in stroke rehabilitation trial methodology and facilitate the search for new treatments [14,17–21].

The scales examined in our study fall into two of the categories described by Lyden and Lau, global scales and activities of daily living (ADL) scales [18]. The MRS and the ISTM are global scales that attempt to quickly group patients into a few very large categories; any change in category must therefore be clinically significant. This approach is intended to maximize interrater reliability, if the borders of each category are well defined. In addition to disability, the MRS and ISTM also capture aspects of handicap, the social and economic consequences of disease. Global scales do have weaknesses: the categories may be vague and no “correct” answer can be objectively determined, scoring is determined by the overall impression of function, and the use of large categories may reduce sensitivity to change.

ADL scales were designed to quantify the amounts of assistance needed to perform specific tasks required for independent living. After using the BI to quantify ADL in stroke patients, Granger and coworkers developed the FIM to address concerns about the BI’s lack of sensitivity to change, ceiling effects, and inadequate assessment of disability in patients with cognitive deficits [22,23]. ADL scales should facilitate detection of smaller changes in disability and for development of explicit scoring guidelines in which the correct answer can easily be determined. The disadvantages of this approach are that the criterion for what constitutes “real” change may be uncertain, that ADL scales have a more complex structure which may decrease interrater reliability, and that they require more time to use.

In this study, we directly compare the sensitivity to change of three scales used in acute stroke trials by studying patients during their stay on an inpatient rehabilitation unit. To study sensitivity to change, we needed a group of subjects likely to have changes in disability. A rehabilitation population is useful because the first few weeks after stroke are a period of rapid clinical improvement and because patients selected for rehabilitation are expected to have a high likelihood of some improvement in disability during treatment [24]. We sought to determine the sensitivity to changes of scales currently used in early intervention trials and to assess the impact on statistical power in designing future rehabilitation treatment trials.

METHODS

Subjects

Consecutive admissions to an academic tertiary care stroke rehabilitation service were included in this prospective study. All patients with ICD-9 (International Classifications of Diseases) cerebrovascular disease diagnoses were allowed, as were patients who also had previous strokes. Patients were excluded from analysis if their primary rehabilitation diagnosis was a condition other than stroke, if data were incomplete, or if rehabilitation was discontinued because of medical or neurological complications. A multidisciplinary team cared for all patients under the supervision of a single attending physician (AWD) during rehabilitation.

Measures

The global scales examined in the study are described in the following:

- Modified Rankin Scale
 - 0 = No symptoms.
 - 1 = No significant disability despite symptoms; able to perform all usual duties and activities.
 - 2 = Slight disability; unable to perform all previous activities but able to look after own affairs without assistance.
 - 3 = Moderate disability; requiring some help, but able to walk without assistance.
 - 4 = Moderately severe disability; unable to walk without assistance and unable to attend to own bodily needs without assistance.
 - 5 = Severe disability; bedridden, incontinent, and requiring constant nursing care and attention.
 - 6 = Dead.

- International Stroke Trial Measure
 - 1 = Alive and fully recovered from stroke.
 - 2 = Alive and independent, but with residual stroke symptoms.
 - 3 = Alive, but dependent on others.
 - 4 = Dead.

The ADL scales examined were as follows:

- Barthel Index. Items rated and weighting:
 - 15 points = walking, transfers.
 - 10 points = feeding, bowel, bladder, toileting, dressing, stairs.
 - 5 points = bathing, grooming.
- Functional Independence Measure. Eighteen items are equally weighted; each is rated from 1 (completely dependent) to 7 (independent without device). The items included feeding, upper-limb dressing, lower-limb dressing, bowel management, bladder management, feeding, bathing, walking, stairs, tub transfer, toilet transfer, bed transfer, comprehension, expression, problem solving, social interaction, and memory.

The UK-TIA (United Kingdom transient ischemic attack) version of the MRS was chosen because published interrater reliability data are available for this version [4,25]. A change of one point was deemed clinically significant because the categories are broad and any change detected would be expected to be large. The ISTM was used as originally published [3]. As with the MRS, any change detected on this scale was considered clinically significant.

The original 100-point version of the BI was used; the guidelines of Collin et al. were followed to clarify the subjective portions of the BI [26]. The BI scores patients on 10 ADL items in increments of five points; a 100-point score indicates that the patient is independent in these 10 items. Following the most stringent recommendations of Collin et al. [26], we chose a 20-point change as being clinically significant. The modified BI of Shah et al. was not included because it has not achieved wide use [27].

The FIM version 4.0 was used according to the standards of the Uniform Data System [28]. While no recommendations exist for what constitutes a clinically significant change on the FIM, a 10-point improvement decreases by almost 50 percent the time required to care for a group of stroke patients in the community [29]. We therefore chose this to be the threshold for change in this study.

Procedures

The rehabilitation team generated FIM scores weekly as part of the clinical routine; members of the team were blinded to this investigation. All raters were trained to the standards of the Uniform Data System for the FIM version 4.0 [28]. The BI scores were derived from the FIM scores based on the scoring algorithm developed by Nyein et al. [30].

The attending physician (AWD) assigned admission and discharge ISTM and MRS scores by using procedures currently in use for acute stroke trials [4]. He received training in use of the MRS and ISTM as a coinvestigator in several acute stroke trials. As part of this training, he met study criteria for test-retest and interrater reliability on these scales. He has also conducted such training on the MRS and BI in eight multicenter stroke trials.

Admission and discharge ISTM and MRS scores were assigned independently based on clinical evaluation of the patient, chart reviews, and FIM scores. MRS scores were assigned in accordance with current acute stroke trial procedures [4].

ISTM scores were assigned as follows: Persons who scored 5 or less on any item on the FIM were rated as a 3 and patients who scored 6 or greater on every FIM item were rated as a 2. An ISTM score of 1 was reserved for patients who scored 7 on all FIM items and whose discharge neurologic exam indicated no stroke-related impairments or return to prestroke neurologic status and functional status. Each patient was then classified as to whether or not a clinically significant change (as defined in the following paragraph) occurred on each scale.

Analyses

Data were analyzed with SAS for Windows, version 6.10. Chi-square analyses using the Cochran-Mantel-Haenszel test were used for pair-wise comparisons of the baseline and discharge scores of each scale. The effect size is a standardized estimate of the power of the test to detect change [31]. The larger the coefficient is, the more responsive the measure is. Standardized response means (SRM) were computed for the FIM and BI as an index of effect size or responsiveness. The SRM is equal to the mean change in scores divided by the standard deviation of the change in scores. The SRM cannot be calculated for ordinal scales with few categories, and therefore this analysis was not performed for the MRS and ISTM. Logistic regression was used to generate receiver operating characteristic (ROC) analyses to evaluate the accuracy of the scales [32].

RESULTS

Ninety-five subjects met entry criteria; all had a significant change on at least one of the scales used. There were 53 women and 42 men. The mean interval from stroke to rehabilitation admission was 9.0 ± 6.4 days, and the length of stay on the rehabilitation service was 19.5 ± 8.3 days. The scores for one patient declined on the BI, FIM, and MRS scales. Review of his chart indicated that he suffered another stroke while on the rehabilitation service but was not transferred. Scores for this individual were included in all analyses. **Table 1** shows the mean and median change scores for each scale. Since most patients did not change on the ISTM, the median change was zero.

The distribution of scores had relatively little shift when patients were evaluated with the global scales, par-

ticularly the ISTM. The ADL scales demonstrated large shifts of patients to higher functional levels. A ceiling effect occurred with the use of the BI, since 25 patients achieved either a 95 or 100; no patient achieved the highest possible score on the FIM. A modest floor effect was also seen with use of the BI because many patients clustered at the lowest possible admission BI scores.

Table 2 shows the number of patients achieving clinically significant changes in one disability scale but not in another. When we compare the global scales, the MRS detected change in 33 patients who were unchanged when measured with the ISTM ($X^2 = 17.56, p < 0.001$). All patients who improved on the ISTM also improved when measured by the MRS. Comparing the two ADL scales, the FIM detected change in 18 patients that the BI did not detect ($X^2 = 9.33, p < 0.001$). The three patients who

Table 1.
Change scores in study population (n = 95).

Measure	Mean Change*	Percent of Subjects Changed† (%)	Median Change‡	Floor Effect, Admission	Ceiling Effect, Discharge
MRS	—	47	1 level (0–2)	17 (18%)	0 (0%)
ISTM	—	24	0 level (0–1)	95 (100%)	0 (0%)
BI	28 ± 16.2	100	30 points (0–70)	5 (5%)	26 (27%)
FIM	23.2 ± 10.6	100	22 points (4–55)	0 (0%)	0 (0%)

*Data are mean \pm standard deviation.
†Percent of subjects with a change of score.
‡Data are medians with ranges in parentheses.
BI = Barthel Index

FIM = Functional Independence Measure
ISTM = International Stroke Trial Measure
MRS = Modified Rankin Scale

Table 2.
Subjects who achieved changes in disability status.*

	Subjects Who DID Change on MRS (n = 55)	Subjects Who DID Change on BI (n = 71)	Subjects Who DID Change on FIM (n = 91)	Subjects Who DID Change on ISTM (n = 23)
Subjects Who Did NOT Change on MRS (n = 40)	—	25 (26%)	37 (39%)	0
Subjects Who Did NOT Change on BI (n = 24)	5 (5%)	—	8 (19%)	4 (4%)
Subjects Who Did NOT Change on FIM (n = 4)	1 (1%)	3 (3%)	—	0
Subjects Who Did NOT Change on ISTM (n = 72)	33 (35%)	52 (55%)	70 (74%)	—

*See Results section of main paper for Chi-square analysis results.
BI = Barthel Index
FIM = Functional Independence Measure
ISTM = International Stroke Trial Measure
MRS = Modified Rankin Scale

achieved clinically significant change on the BI but not on the FIM had improvements on the same tasks on each scale, but the FIM score changes were 1 or 2 points while the corresponding BI changes were 5 or 10 points.

Comparisons between the global and ADL scales were also informative. The BI detected change in 25 more patients than the MRS ($X^2 = 9.92$, $p < 0.002$) and in 52 patients more than the ISTM. Chi-square analysis comparing the BI and the ISTM was not possible because of the small numbers of subjects in some cells. However, the BI once again displayed a ceiling effect: of the five patients who changed on the MRS but not the BI, all began rehabilitation with a BI score of 95 and thus could not improve the 20 points required for a clinically significant change. Similarly, the ISTM detected a significant change in four patients who did not achieve a clinically significant change on the BI because of high initial BI scores. The FIM detected change in a significantly larger number of patients than the MRS ($X^2 = 8.07$, $p < 0.005$); Chi-square analysis comparing the FIM and ISTM was again impossible.

Standardized response means were calculated for the BI and the FIM. The FIM (SRM = 2.18) was more responsive than the BI (SRM = 1.72). The larger coefficient means that the FIM is more responsive to change than the BI. These analyses cannot be performed on the MRS and ISTM because they are ordinal scales.

Based on the Chi-square and effect size analyses, the FIM was selected as the reference measure for the ROC analysis. Sensitivity, specificity, and the area under the ROC curves were computed for the BI, MRS, and ISTM. The curves for each measure are presented in the **Figure**. The BI has the highest predictive accuracy as indicated graphically by the quick rise of the curve as well as by the C-statistic ($C = 0.82$), which measures area under the curve. The MRS and ISTM have curves that do not rise as quickly; C-statistics for the MRS and ISTM are 0.59 and 0.51, respectively. The BI has higher sensitivity and specificity than the MRS and ISTM when compared to the FIM.

DISCUSSION

Disability scales are the current end points for acute stroke treatment trials. Because some of these intervention trials (i.e., t-PA) have been successful in improving outcome, stroke rehabilitation studies may be expected to employ similar instruments. We report the relative sensi-

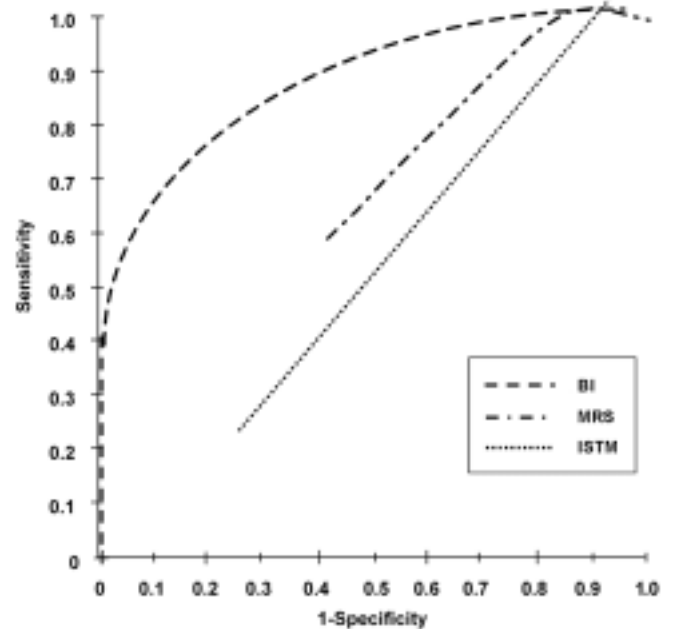


Figure. ROC analysis. Sensitivity and specificity of BI, MRS, and ISTM compared to FIM bibliography.

tivities to change of four disability scales in a stroke rehabilitation population to help investigators evaluate the use of these measures in the postacute treatment setting. We found large differences in the sensitivity to change among the scales examined.

The criterion for “clinically significant” change chosen for each scale deserves comment. We followed the advice of Adams to “avoid inflation of minimal changes that are of dubious clinical significance” [17]. The choice of a 1-point change on the 5-point MRS or the 4-point ISTM is reasonable because inspection of the scales reveals that the categories are quite large; any change detected by these measures must be large and presumably clinically significant. The absolute amount of change defined as clinically significant for the ADL scales was deliberately smaller than that chosen for the global scales, thus exploiting a major strength of the ADL scales. Nonetheless, we believe the levels we set for the two ADL scales detect a change clearly meaningful to the patient, family, and clinician. Literature has had little explicit comment about what constitutes a “real” change in the BI. Wade et. al. and Shah both suggest that a 20-point threshold would certainly indicate important change [5,27]. This criterion may be overly rigorous; even so, the BI detected change in a much larger proportion of patients

that the MRS. For the purposes of a treatment trial, a smaller difference between the treatment and control groups on the BI might be adequate.

There are few published comparisons of the FIM and BI in stroke rehabilitation populations. The FIM evolved from the BI and was intended to address concerns about the BI, including lack of sensitivity to change, ceiling effects, and inadequate assessment of disability in patients with primarily cognitive deficits. Our data clearly demonstrate that the FIM is more responsive to change and less prone to ceiling effects than the BI. A 10-point change in a FIM score requires net improvement of one level or more in many items. Based on a recent study [29], a 10-point FIM score improvement would reduce by nearly 50 percent the average amount of time required to physically care for this cohort of patients in the home, a change which patient and family would surely consider significant.

Although the ceiling effect of the BI is well known, the magnitude of the effect in our study was unexpected. Twenty percent of our subjects achieved the maximum 95 or 100 score on the BI, while no patient achieved the maximum score on the FIM. Thus, the BI classified some of our patients as "normal" when they continued to have measurable disability when evaluated with the FIM. A floor effect for the BI was also present, suggesting that the BI would also fail to detect differences in the most severely affected population. Our findings differ from those reported by van der Putten et al [33], who reported no significant floor or ceiling effects for either the BI or FIM; the differences observed in our study may be related to the differences between United States and United Kingdom rehabilitation settings.

Can results obtained with our patients be generalized to the cohort of patients used in all acute and postacute stroke trials? The population used in our study was not a random sample of all stroke patients or was it those who necessarily represent the population that would be enrolled in all intervention trials. By the nature of a rehabilitation service, those patients with immediate and complete recovery after stroke are excluded, as are those who die or who are severely disabled. Our patients were mostly made up of the moderately disabled, the "middle band" of stroke patients [24]. Since most stroke intervention trials also attempt to target this group by excluding mildly and severely affected subjects, we argue that our cohort is actually enriched in the population of greatest interest to investigators. For example, in the National

Institutes of Neurological Disorders and Stroke (NINDS) t-PA trial [34], 66 percent of survivors in the placebo arm scored between 2 and 5 on the MRS and would potentially be included in our study population.

Our data support the use of ADL scales over the use of global scales in all stroke intervention trials. The use of common measures would allow comparisons across the full spectrum of treatment strategies, including rehabilitation. This will be particularly helpful in cost-benefit analyses that compare pharmaceutical treatments with behavioral interventions. If early and late treatments are combined, a common end point will be necessary. Those who select a global scale may be sacrificing the ability to detect change in disability on the assumption that they will be compensated by very high interrater reliability. While there are studies reporting adequate interrater reliability of the MRS in tightly controlled circumstances, we are not aware of a study that directly compares the interrater reliability of the BI, the ISTM, and the MRS under conditions that realistically simulate a multicenter trial. One study (using stroke research nurses who worked together at a single center) actually found the BI to have better interrater reliability than the MRS [6].

When an ADL scale as an efficacy measure in stroke trials is chosen, the FIM appears to be methodologically superior to the BI, at least in U.S. stroke rehabilitation settings. The FIM detected change in more patients than the BI and did not exhibit the ceiling and floor effects seen in the BI. While the FIM is more complex and takes longer to learn and administer, it is designed for use by nonphysicians. The small additional cost of a therapist's time may be offset by the need to enroll fewer patients when using a more sensitive measure. Scoring of the FIM via telephone interview is used widely and can greatly reduce the cost of collecting outcome data [25].

Our results must be evaluated in the context of the recent NINDS t-PA trial in which subjects were classified into "favorable" and "unfavorable" outcome groups [34]. First, creation of a dichotomous outcome measure may be appropriate for evaluating a thrombolytic agent where the effect of treatment is expected to be either quite dramatic or nonexistent. This approach may not be optimal for other classes of agents in which the treatment effect may be less dramatic [14]. Incremental improvement or worsening in function may be missed by an unresponsive outcome measure, and agents may not be correctly evaluated for efficacy or safety. Second, using dichotomous outcome measures may increase the importance of the ceiling

effects present in currently used outcome measures. Unwary clinicians or health care planners may interpret “favorable” to mean “cured.” The false classification of patients as “cured” based on a measure with a large ceiling effect, such as the BI, may overstate the effect of a treatment.

CONCLUSION

In summary, for this moderately disabled stroke population, we found wide variations in the sensitivity to change in three disability measures currently used in stroke intervention trials and one disability measure widely used in rehabilitation centers. We emphasize that sensitivity to change is only one factor in choosing a disability scale and that reliability, validity, specificity, and practicality also must be considered. Further direct comparison of the properties of disability scales may be useful in optimizing stroke trial design.

ACKNOWLEDGMENT

We thank Cynthia L. Reynolds, RN; Phyllis Schuessler, RPT; and the members of the Stroke Management and Rehabilitation Team for their assistance in data collection.

REFERENCES

- Rankin J. Cerebral vascular accidents in patients over the age of 60: II. Prognosis. *Scott Med J* 1957;2:200–15.
- Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965;14:61–65.
- Lindley R. Very simple methods of assessing outcome after stroke can be valid. *Cerebrovasc Dis* 1992;2:209.
- van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJA, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988;19:604–7.
- Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability. *International Disability Studies* 1988;10:64–67.
- Wolfe CDA, Taub NA, Woodrow EJ, Burney PGA. Assessment of scales of disability and handicap for stroke patients. *Stroke* 1991;22:1242–44.
- 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.
- Shah S. In praise of the biometric and psychometric qualities of the Barthel Index. *Physiotherapy* 1994;80:769–71.
- van Gijn J, Warlow CP. Down with stroke scales! *Cerebrovasc Dis* 1992;2:244–46.
- Wood-Dauphinee S, Williams JI, Shapiro SH. Examining outcome measures in a clinical study of stroke. *Stroke* 1990;21:731–39.
- Gresham GE, Phillips TF, Labi MLC. ADL status in stroke: relative merits of three standard indexes. *Am J Phys Med Rehabil* 1980;61:355–58.
- Keith RA, Granger CV, Hamilton BB, Sherwin FS. The Functional Independence Measure: a new tool for rehabilitation. In: Eisenberg MG, Gresiak RC, editors. *Advances Clinical Rehabilitation*. New York: Springer; 1987.
- Deyo RA, Inui TS. Toward clinical application of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res* 1984;19:275–89.
- Dorman PJ, Sandercock PAG. Considerations in the design of clinical trials of neuroprotective therapy in acute stroke. *Stroke* 1996;27:1507–15.
- Hallenback JM, Frerichs KU. Stroke therapy: it may be time for an integrated approach. *Arch Neurol* 1993;50:768–70.
- Adams HP, Biller J. Ischemic cerebrovascular disease. In: Bradley WG, Daroff RB, Fenichel GM, Marsden CD, editors. *Neurology in clinical practice*. 2nd ed. Boston: Butterworth-Heinemann; 1996. p. 993–1031.
- Adams HP. Trials of trials in acute stroke. *Stroke* 1993;24:1410–15.
- Lyden PD, Lau GT. A critical appraisal of stroke evaluation and rating scales. *Stroke* 1991;22:1345–52.
- Cote R, Battista RN, Wolfson CM, Hachinski V. Stroke assessment scales: guidelines for development, validation, and reliability assessment. *Can J Neurol Sci* 1988;15:261–65.
- Orgogozo JM, Dartigues JF. Methodology of clinical trials in acute cerebral ischemia. *Cerebrovasc Dis* 1991;1(Suppl):100–11.
- Edwards DA, Chen Y, Diringer MN. Stroke scales are reliable and valid for ischemic and hemorrhagic stroke. *Stroke* 1995;26:1852–58.
- Granger CV, Hamilton BB, Gresham GE. The stroke rehabilitation outcome study: Part I. General description. *Arch Phys Med Rehabil* 1988;69:506–9.
- Granger CV, Hamilton BB, Gresham GE, Kramer AA. The stroke rehabilitation outcome study: Part II. Relative merits of the total Barthel Index score and a four-item subscore in predicting patient outcomes. *Arch Phys Med Rehabil* 1989;70:100–3.
- Garraway M. Stroke rehabilitation units: concepts, evaluation and unresolved issues. *Stroke* 1985;16:178–81.

25. UK-TIA Study Group. The UK-TIA aspirin trial: Interim results. *Br Med J* 1988;296:316–20.
26. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. *Int Disabil Stud* 1988;10:61–63.
27. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989;42:703–9.
28. Uniform Data System for Medical Rehabilitation. Guide for the Uniform Data Set for Medical Rehabilitation [abstract]. 1993.
29. Granger CV, Cotter AC, Hamilton BB, Fiedler RC. Functional assessment scales: a study of persons after stroke. *Arch Phys Med Rehabil* 1993;74:133–38.
30. Nyein K, McMichael L, Turner-Stokes L. Can a Barthel score be derived from the FIM? *Clin Rehabil* 1999;13:56–63.
31. Guyatt GH, Kirschner B, Jaeschke R. Measuring health status: what are the necessary measurement processes? *J Clin Epidemiol* 1992;45:1341–45.
32. Hanley JA. Receiver operating characteristic methodology: the state of the art. *Crit Rev Diagn Imaging* 1989;29:307–34.
33. van der Putten JJMF, Hobart JC, Freeman JA, Thompson AJ. Measuring change in disability after in-patient rehabilitation; Comparison of the responsiveness of the Barthel Index and Functional Independence Measure. *J Neurol Neurosurg Psychiatry* 1999;66:480–84.
34. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581–87.

Submitted for publication June 26, 2002. Accepted in revised form October 11, 2002.