

Measurements of Acute Cerebral Infarction: A Clinical Examination Scale

Thomas Brott, MD, Harold P. Adams Jr., MD, Charles P. Olinger, MD,
John R. Marler, MD, William G. Barsan, MD, José Biller, MD, Judith Spilker, RN,
Renée Holleran, RN, Robert Eberle, Vicki Hertzberg, PhD, Marvin Rorick, MD,
Charles J. Moomaw, PhD, and Michael Walker, MD

We designed a 15-item neurologic examination stroke scale for use in acute stroke therapy trials. In a study of 24 stroke patients, interrater reliability for the scale was found to be high (mean $\kappa=0.69$), and test-retest reliability was also high (mean $\kappa=0.66-0.77$). Test-retest reliability did not differ significantly among a neurologist, a neurology house officer, a neurology nurse, or an emergency department nurse. The stroke scale validity was assessed by comparing the scale scores obtained prospectively on 65 acute stroke patients to the patients' infarction size as measured by computed tomography scan at 1 week and to the patients' clinical outcome as determined at 3 months. These correlations (scale-lesion size $r=0.68$, scale-outcome $r=0.79$) suggested acceptable examination and scale validity. Of the 15 test items, the most interrater reliable item (pupillary response) had low validity. Less reliable items such as upper or lower extremity motor function were more valid. We discuss methods for improving the reliability and validity of brief examination scales to be used in stroke therapy trials. (*Stroke* 1989;20:864-870)

The introduction of new therapies for acute stroke has renewed interest in clinical measurements of the severity of cerebral infarction. Traditional measures such as mortality or long-term functional status are not well suited to trials of acute therapy where immediate effects must be emphasized. Limited clinical measurements of patients with acute cerebral infarction have been used in previous therapy outcome studies.¹⁻⁴ Later studies employed neurologic rating scales, but these scales were not analyzed for length of examination time, interexaminer reliability, intraexaminer reliability, or correlation to the extent or location of the vascular lesion.⁵⁻¹⁰ More recently, the issue of reliability of the clinical assessment has

been addressed,¹¹⁻¹⁵ but clinimetric validity of the assessment tools has not been studied.

We developed a system for examination of patients with acute cerebral infarction in three phases. First, we derived a brief but comprehensive scale of the neurologic examination. Then the examination and scale were evaluated for test-retest reliability and interrater reliability. Second, the examination and scale were "field-tested" as part of a prospective therapy (naloxone) study of 65 patients with acute cerebral infarction. Finally, validity was assessed by correlating the scale scores with two other measures of stroke severity: computed tomography (CT) measurement of volume of infarction and clinical outcome at 3 months.

Subjects and Methods

Developing the Stroke Scale

For design of the examination format, a pilot study consisting of 10 stroke patients was performed with two previously used assessment forms (Toronto Stroke Scale, Oxbury Initial Severity Scale),^{16,17} and a third scale (Cincinnati Stroke Scale, which graded speech, drift of the affected arm, drift of the affected leg, and grip strength). Patients were tested within 3 weeks of cerebral infarction. Results were analyzed qualitatively, and a composite examination scale was derived using items from each of the three scales. Included were

From the Departments of Neurology (T.B., C.P.O., J.S., R.E., M.R.), Emergency Medicine (W.G.B., R.H.), and Environmental Health (V.H., C.J.M.), University of Cincinnati, Cincinnati, Ohio; the Department of Neurology (H.P.A., J.B.), University of Iowa, Iowa City, Iowa; and the Division of Stroke and Trauma (J.R.M., M.W.), National Institute of Neurological Disorders and Stroke, Bethesda, Maryland.

Supported by United States Public Health Service, National Institutes of Health, National Institute of Neurological Disorders and Stroke Contracts N01-NS-2324, TO 1 (University of Cincinnati) and N01-NS-2326, TO 1 (University of Iowa).

Address for reprints: Thomas Brott, MD, Department of Neurology, University of Cincinnati Medical Center, 231 Bethesda Avenue, Cincinnati, OH 45267-0525.

Received November 7, 1988; accepted January 9, 1989.

Stroke Scale

1.a. Level of Consciousness	Alert Drowsy Stuporous Coma	0 1 2 3	
1.b. LOC Questions	Answers both correctly Answers one correctly Incorrect	0 1 2	
1.c. LOC Commands	Obeys both correctly Obeys one correctly Incorrect	0 1 2	
2. Pupillary Response	Both reactive One reactive Neither reactive	0 1 2	
3. Best Gaze	Normal Partial gaze palsy Forced deviation	0 1 2	
4. Best Visual	No visual loss Partial hemianopia Complete hemianopia	0 1 2	
5. Facial Palsy	Normal Minor Partial Complete	0 1 2 3	
6. Best Motor Arm	No drift Drift Can't resist gravity No effort against gravity	0 1 2 3	
7. Best Motor Leg	No drift Drift Can't resist gravity No effort against gravity	0 1 2 3	
8. Plantar Reflex	Normal Equivocal Extensor Bilateral extensor	0 1 2 3	
9. Limb Ataxia	Absent Present in upper or lower Present in both	0 1 2	
10. Sensory	Normal Partial loss Dense loss	0 1 2	
11. Neglect	No neglect Partial neglect Complete neglect	0 1 2	
12. Dysarthria	Normal articulation Mild to moderate dysarthria Near unintelligible or worse	0 1 2	
13. Best Language	No aphasia Mild to moderate aphasia Severe aphasia Mute	0 1 2 3	
14. Change from Previous Exam	Same Better Worse	S B W	
15. Change from Baseline	Same Better Worse	S B W	

FIGURE 1. Item-by-item testing procedure and grading are defined in stroke scale glossary, which is available upon written request.

items to test for the presence of neurologic signs in the distribution of each of the major arteries of the brain (e.g., visual field assessment for infarction in the distribution of the posterior cerebral artery). Mental status assessment was supplemented by adding two items from the Edinburgh-2 Coma Scale.¹⁸ Discussion with other investigators participating in the National Institute of Neurological Disorders and Stroke (NINDS) stroke treatment studies resulted in the addition of categories for sensory function, pupillary response, and plantar response. The final examination format is shown in

Figure 1. A detailed glossary was provided to standardize the bedside examination procedures.

Each item of the neurologic examination was assessed and rated independently, and no weighting of the items was used. This format allowed for detection of changes during sequential examinations for each component. The numbers of grades for each component varied, ranging from three grades for the measurement of pupillary responses to four grades for the limb motor examinations (Figure 1). For all parameters, a value of 0 was normal. The differences between the grades were not subtle. For example, motor performance for the upper extremity was graded as 0=normal, 1=drift (with arms outstretched), 2=inability to resist gravity for 10 seconds (with arms outstretched), and 3=no effort against gravity. The new scale was also designed to be done easily and quickly at the bedside so that sequential examinations could be performed practically. It was anticipated that some components of the scale might not be measurable in all patients. For example, limb ataxia might not be detectable in a paralyzed limb or neglect in an obtunded patient.

To compare the results of the scale with the results of the CT examinations and clinical outcome at 3 months, the scores of all the scale items were added to give a composite score. No weighting between individual parameters was used.

Assessing Stroke Scale Reliability

Twenty-four patients with acute cerebral infarction (onset within 1 week) were examined twice within a 24-hour interval. Each examination was carried out by a staff neurologist while the other three examination team members (neurology house officer, neurology nurse-clinician, and emergency department nurse-clinician) observed; each team member independently performed scoring. At the time of the second examination, team members had no record of their own previous examination and no knowledge of the scores assigned by the other examiners. This two-examination method was chosen rather than eight separate examinations to minimize patient fatigue and potential practice effect.

The scores were tabulated, and an item-by-item analysis was performed. Reliability was assessed in regard to agreement from the first examination to the second (test-retest reliability) and in regard to agreement among the examiners (interrater reliability). The results were analyzed using the κ statistic.¹⁹

Use of the Stroke Scale in a Therapy Trial

At the University of Cincinnati and the University of Iowa in 1984 and 1985, prospective field-testing of the scale was performed as part of a study of naloxone as early treatment for acute cerebral infarction. All patients were treated within 48 hours of stroke onset or later progression (mean=13 hours, 26 minutes; SD=10 hours, 11 minutes). Practice examinations were completed by the investigators before actual treatment study was begun. During

the study, 65 patients were examined before therapy, five times during a 24-hour treatment period and once at 7 days and at 3 months. Each patient was first examined within 48 hours of onset of an acute, nonhemorrhagic, cerebral infarction (mean time=10 hours, 36 minutes). Often the serial examinations for a given patient were performed by more than one investigator. When an individual examination item could not be evaluated (e.g., sensory function in a mute patient), that item was scored as nontestable, and a number value was not assigned. Of the 65 cerebral infarctions, 23 involved the right hemisphere, 37 involved the left hemisphere, and five involved the brainstem or cerebellum. At the time of admission, the infarction was thought to be large-artery atherothrombotic in 52%, small-artery atherothrombotic in 18%, cardiac embolic in 17%, and large-artery atheroembolic in 12%.

Methods for Assessing Stroke Scale Validity

To test the validity of our stroke scale, we correlated the scores to volume of infarction as measured by CT. As we anticipated that neither the scale scores nor the CT lesion volumes would be normally distributed, the Spearman's rank-order method was used.²⁰ For the stroke scale-infarct volume correlations, the 65 naloxone patients were serially examined by CT scan without contrast at admission (before treatment), with contrast (if possible) at 7–10 days after stroke, and at 3 months (the methods for the CT scan lesion volume measurements are the subject of another report in this issue of *Stroke*²¹). Spearman's rank-order correlation coefficients were calculated comparing the neurologic examinations at pretreatment baseline and 7 days to the CT scans performed at baseline, 7 days, and 3 months.

We also correlated the scale scores to three different measures of functional outcome, as determined at 3 months. For the correlation of our stroke scale to patient outcome, the 3-month outcome measures were expressed as a numerical outcome score, modified from the scales suggested by Spence and Donner.²² For the patient's performance class, 1 point was scored for independence, 2 points for mild impairment, 3 points for moderate impairment, and 4 points for severe impairment. For the patient's placement class, 1 point was scored for no limitation, 2 points for mild limitation, 3 points for moderate limitation, and 4 points for severe limitation. For the patient's location in the community, 1 point was scored for returning home, 2 points for residing in a minimal care facility, and 3 points for residing in a skilled nursing facility, a hospital, or a rehabilitation facility. A total outcome score was computed for each patient. Spearman's rank-order correlation coefficients were then computed comparing the 3-month outcome scores with the stroke scores obtained for each patient at baseline and 7 days.

TABLE 1. Stroke Scale Reliability by κ Analysis

Examination item	Interexaminer reliability examinations		
	First	Second	Both
Pupillary response	1.00*	0.90*	0.95*
Best motor arm	0.79	0.89	0.85
Best motor leg	0.81	0.84	0.83
Best gaze	0.82	0.84	0.82
Best visual	0.89	0.84	0.81
LOC questions	0.96	0.90	0.80
Plantar reflex	0.59	0.65	0.67
Best language	0.61	0.76	0.64
Sensory	0.63	0.74	0.60
LOC commands	1.00	0.92	0.58
Neglect	0.54	0.56	0.58
Limb ataxia	0.52	0.84	0.57
Facial palsy	0.46	0.59	0.57
Dysarthria	0.44	0.56	0.55
LOC	0.42	0.670	0.49

LOC, level of consciousness.

*Expressed as mean κ statistic of each pair of four observers. κ statistic quantitatively indicates amount of agreement that exists beyond chance. Perfect agreement would=1.00, and perfect disagreement would=-1.00. For human rater agreement, >0.50 is thought to be acceptable and >0.80 is thought to be excellent. Adapted from Fleiss¹⁹ with permission.

The individual scale items were also analyzed as to whether they were scored normal when the patient had a clinical neurologic deficit. For example, the pupillary response could be normal despite aphasia and hemiplegia. The individual items were also analyzed as to whether they did or did not change as the patient's overall clinical status changed.

Results

The examination could be performed quickly, requiring a mean of 6.6 ± 1.3 minutes for the 48 examinations performed to assess interexaminer and intraexaminer reliability. In the naloxone therapy trial, the scale was successfully completed in all 65 patients regardless of the stroke severity or the time of day (many examinations were performed between midnight and 6 AM). For the admission examination of an individual patient, a mean of 1.3 items per patient could not be evaluated out of the 15 items in each examination (e.g., limb ataxia was untestable in patients with total hemiplegia).

Interrater agreement using a κ analysis was high among the four examiners (Table 1). With perfect agreement=1.00 and perfect disagreement=-1.00, the mean κ for the 15 items was 0.69. Agreement was particularly high for six items ($\kappa \geq 0.80$): pupillary response (0.95), best motor arm performance (0.85), best motor leg performance (0.83), best gaze (0.82), and level of consciousness (LOC) questions (0.80). The lowest agreement was for the qualitative assessment of level of consciousness ($\kappa=0.49$).

TABLE 2. Intraexaminer Agreement for Two Examinations of 24 Patients

Examiner	Mean κ
MD A (neurologist)	0.77*
MD B (neurology house officer)	0.70
RN A (neurology nurse clinician)	0.68
RN B (emergency dept. nurse clinician)	0.66*

*Difference between agreement of MD A and RN B not statistically significant.

Test-retest reliability for a given examiner was also excellent (Table 2). Scoring by the neurologist was the most consistent between examinations ($\kappa=0.77$). However, scoring consistency was only slightly lower for the other examiners (Table 2).

The correlation between the first examination scores and the second examination scores was 0.98 ($p<0.0001$). For a given patient, the correlation of one examiner's score for the first exam with a different examiner's score for the second examination was very high; for example, a first examination by the neurologist of an individual patient correlated with a second examination of that patient by the emergency department nurse with Spearman's correlation=0.98.

The Spearman's correlation between the total stroke scale score at 7 days and the CT scan lesion volume at 7 days was 0.74 (Table 3). The patients' initial neurologic deficit as measured by the scale also correlated with the 7–10-day CT lesion volume (correlation coefficient=0.78). The scale-CT correlation at 7 days for patients with left hemisphere infarctions was 0.72, while that correlation for patients with right hemisphere infarctions was 0.74.

An abbreviated examination score was calculated using the three items from the examination that were the most reliable and that were estimated to reflect clinical change closely (LOC questions, best motor arm, best motor leg). This abbreviated examination at 7 days correlated less well with CT lesion volume at 7–10 days (correlation coefficient=0.68) than did the 7-day full scale scores (0.79).

The stroke scale scores at 7 days corresponded to eventual patient clinical outcome with a Spearman's correlation of 0.71 ($p=0.0001$) (Table 4). Correlation was equally high for the 7-day stroke scale to either the patient performance class or to placement class and lower for the scale to the patient location in the community (Table 4). The admission stroke scale scores also correlated with

TABLE 3. Stroke Scale Score Compared With Computed Tomographic Scan Lesion Size

Comparison	Correlation coefficient	p
Baseline score vs. baseline lesion volume	0.39	0.0033
Baseline score vs. 7-day lesion volume	0.78	0.0001
7-day score vs. 7-day lesion volume	0.74	0.0001

TABLE 4. Stroke Scale Score Compared With Patient Outcome

Outcome measure (at 3 mo.)	Admission scale score*	7-day scale score
Performance class	0.59†	0.75†
Placement class	0.62†	0.76†
Location in community	0.19	0.33†
Total outcome score	0.53†	0.71†

*Numbers in columns represent Spearman's correlation coefficients of stroke scale scores with outcome measures.

† $p=0.0001$.

the eventual patient outcome, with a Spearman's correlation of 0.53 ($p=0.0001$).

For eight of the 15 examination items, more than 50% of the patients were scored as normal at the time of admission (Table 5). In other words, these items were not sensitive to the occurrence of an acute stroke. For pupillary response, 94% of the patients were normal; for level of consciousness (qualitative), 90% of the patients were normal. For any examination item, a spread of scores in the abnormal range would be preferable, allowing for detection of the stroke and for differentiation of minor strokes from major strokes. A spread of scores from slightly abnormal to grossly abnormal was best achieved by the items best motor arm and best motor leg.

An examination scale will also be more valid if the subject's scores change whenever the subject's condition (or performance) changes. Most of our patients improved clinically. However, four of the 15 examination items changed minimally ($\leq 10\%$): facial palsy, plantar reflex, dysarthria, and language (Table 5). Change in an item score, as it relates to lesion size, may have been overstated by the items limb ataxia (59% improvement) and best gaze (52% improvement). The remaining 10 examination items changed an average of 25% over 7 days. The change for 13 of the 15 items reflected improvement.

The scale also contains a requirement that the examiner conclude qualitatively for each examination whether the patient has changed neurologically from baseline and whether the patient has changed neurologically from the previous examination. The scoring choices are "Same," "Better," or "Worse." To compare this qualitative method with the scale, we defined "Same" (for the scale) to be a change of 0–1 scale point, "Better" to be an improvement of 2 or more scale points, and "Worse" to be a deterioration of 2 or more scale points. With these criteria, comparing the quantitative criteria for patient change with the investigator's qualitative impression of patient change (from baseline to 7–10 days) we found agreement for 40 of the 63 patients surviving at 7–10 days (63%).

Discussion

For a clinical stroke scale to be useful in the acute care setting it should be simple, valid, and reliable. These qualities are particularly important given the

TABLE 5. Stroke Scale Item Analysis*

Examination item	% Patients normal on admission (n=65)	% Patients not testable on admission† (n=65)	% Improvement for item score at 7 days (n=63)
Level of consciousness (LOC)	90	0	17‡
LOC questions	71	8	41
LOC commands	76	8	27
Pupillary response	94	2	33
Best gaze	63	0	52
Best visual	54	17	43
Facial palsy	24	0	-2
Best motor arm	21	0	14
Best motor leg	35	0	32
Plantar reflex	30	5	7
Limb ataxia	35	46	59
Sensory	44	10	-15
Neglect	62	10	30
Dysarthria	41	11	-1
Best language	57	0	6

*Analysis from 7-day examination of 65 patients in naloxone therapy trial.

†Not testable refers to inability to assess for given item because of other deficits (e.g., patients with right hemiplegia could not be assessed for right-sided appendicular ataxia).

‡% improvement reflects change only for those patients not normal for that item at first examination.

necessity for a multicenter design to recruit an adequate number of patients for future stroke therapy trials.

Our scale could be completed in 6.6 minutes in acutely ill patients and could be administered consistently by nurses or physicians. The brevity minimized patient fatigue and allowed serial examination over the initial hours after therapeutic intervention. While concise, the scale covered most neurologic signs and allowed identification of clinical change or potential response to therapy. Unfortunately, no bedside examination that can be done in a reasonable time can adequately measure subtle or complex behavioral signs.

Our scale contains 15 examination items and 51 raw score choices. We favor brevity and restricted scoring choices for the ideal examination. Unfortunately, gains in reliability achieved by shortening an assessment tool must be weighed against the loss in examination accuracy as the assessment tool becomes less comprehensive.²³ For example, our scale results might become more consistent by eliminating assessment of the visual fields (which is frequently difficult in acutely ill patients); however, we would lose the only examination item likely to detect occipital infarction. An abbreviated form of our scale (level of consciousness questions, motor arm, and motor leg) was more reliable than the total scale, but it was less valid in that the scores correlated less well with infarction size as indicated by CT scan. Any abbreviated scale has less capability of testing for the wide varieties of clinical ischemic syndromes. Any future trial of a treatment in patients with acute or progressing cerebral infar-

tion must include a sufficient number of examination parameters to assess the wide gamut of neurologic presentations.

Our scale provides consistency (i.e., clinimetric reliability) (Tables 1, 2) with its κ coefficients being uniformly in the good to excellent range.¹⁹ They compare favorably with the κ analysis of the Canadian Neurological Scale¹⁵ and exceed the reliability performance reported for the Stroke Data Bank examination scale.¹⁴ The least reliable item, qualitative assessment of the level of consciousness, is supplemented by two items from the modified Edinburgh-2 Coma Scale that have acceptable to excellent reliability (Table 1).

To measure the reliability of this scale we used the κ measure of agreement, which is more rigorous than the Pearson correlation method.¹⁴ For a coefficient of 1.00, the κ method requires that two examiners agree completely. Any disagreement, however slight, drops the coefficient value. For the Pearson correlation method, the coefficient may be very high as long as any disagreements are consistently related and slight. For example, if one examiner's ratings usually disagree with a second examiner's ratings, the κ coefficient will be low regardless of the direction in which the scores may differ; the Pearson correlation could be high if the first rater's scores are consistently higher (or lower) than the second rater's scores.

Reliability results were probably improved by having one examiner and three observers together scoring each patient on two occasions (albeit scoring independently and without knowledge of the other scorer's opinions) rather than having each

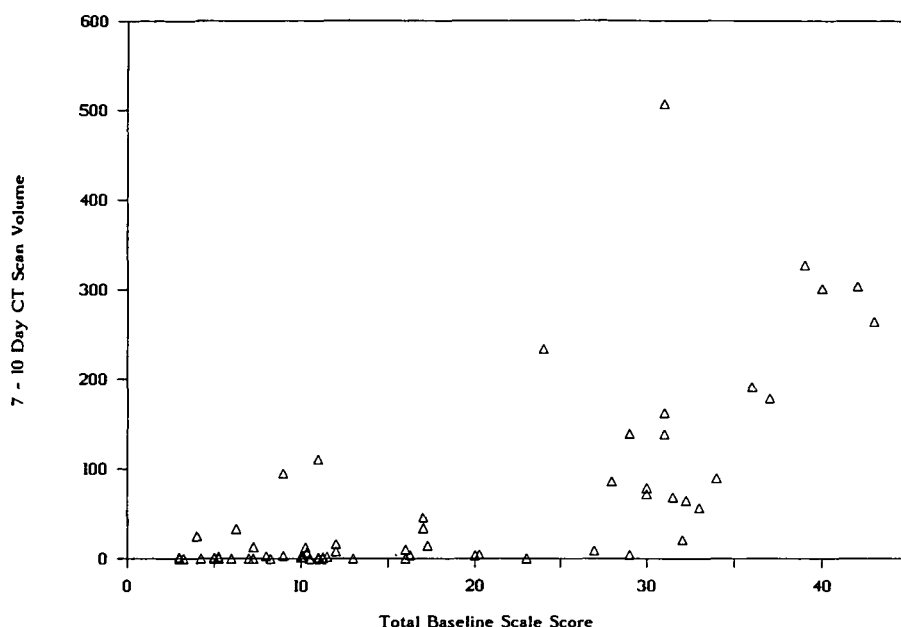


FIGURE 2. The stroke scale score of each patient at time of admission is plotted against that patient's infarction volume as measured from 7-10-day CT scan.

examination team member by himself examining each patient on two occasions. This technique means that changes in patient performance induced by fatigue or practice effect are less likely to cause differences between the examinations. Reliability might have been improved in that the examiners' scores for the second examination may have been influenced by recollections of the first examination (though examiners were blinded to the first examination scores). Reliability may also be anticipated to be lower for other examiners, as the examination team members for this study developed the scale and so became experienced together before the reliability testing.

The three general components of clinimetric validity are reliability, accuracy, and suitability.²³ Accuracy of a scale is best determined by comparing the scale to a definitive standard. For most clinical scales in disease (e.g., Apgar score, Glasgow Coma Scale) a definitive standard is not available. Accuracy is then assessed by comparison of the scale to another reference procedure(s) assumed related to what the scale purports to measure.^{23,24} Using CT infarction volume and patient outcome at 3 months as the reference procedures, we found our stroke scale to be accurate (Tables 3, 4). The correlations between the scale scores and the CT lesion volumes and patient outcome were numerically strong. However, as illustrated by Figure 2, the correlations should be interpreted cautiously given the clustering of patients with normal CT scans at 7-10 days and the scatter of the remaining data points.

The suitability aspect of clinimetric validity refers to the qualitative "common sense value" of a scale for its intended purpose.²³ For acutely ill stroke patients, a reliable and valid scale requiring written responses or requiring 60 minutes for completion would not be suitable. Our stroke scale is brief, it

can be performed by physicians or nurses, it can be repeated easily, it can be performed in the middle of the night with paralyzed or mute patients, and it requires very little training.

Improving our or any other stroke assessment scale with respect to both reliability and validity may be difficult. For example, in our study the pupillary response to light was the most reliable item in the 15-item scale but was rated normal for 94% of the patients, did not change as the patients changed clinically, and did not correlate with the CT scan anatomic measure of lesion size. In comparison, the scale item for upper extremity motor function was less reliable, but the scores changed in parallel to changes in the patient, and the scores correlated to the CT scan lesion size. In future studies, item-by-item correlation from scales such as this one to other measures (e.g., CT scan lesion size, patient outcome at 3 months) could allow elimination of low-sensitivity items and fine-tuning of the high-sensitivity items.

Note added in proof. Following the completion of this study, an independent evaluation of the authors' examination scale was carried out by Goldstein et al.²⁵

Acknowledgments

The authors wish to thank Drs. Carlos Kase, Oscar Reinmuth, and Percy Karanjia for their assistance in design of our stroke scale.

References

1. Baker RN, Broward JA, Fang HC, Fisher CM, Groch SN, Heyman A, Karp HR, McDevitt E, Scheinberg P, Schwartz W, Toole JF: Anticoagulant therapy in cerebral infarction. *Neurology* 1962;12:823-835
2. Fisher CM: Anticoagulant therapy in cerebral thrombosis and cerebral embolism. *Neurology* 1961;11:119-131

3. Siekert RG, Whisnant JP, Millikan CH: Surgical and anticoagulant therapy of occlusive cerebrovascular disease. *Ann Intern Med* 1963;58:637-641
4. Fisher CM: The use of anti-coagulants in cerebral thrombosis. *Neurology* 1958;8:311-332
5. Patten BM, Mendell B, Curtin W, Carter S: Double-blind study of the effects of dexamethasone on acute stroke. *Neurology* 1972;22:377-383
6. Mathew NT, Rivera VM, Meyer JS, Charney JZ: Double-blind evaluation of glycerol therapy in acute cerebral infarction. *Lancet* 1972;2:1327-1329
7. Norris JS: Steroid therapy in acute cerebral infarction. *Arch Neurol* 1976;33:69-71
8. Bauer RB, Tellez H: Dexamethasone as treatment in cerebrovascular disease 2. A controlled study in acute cerebral infarction. *Stroke* 1973;4:547-555
9. Kaste M, Fogelholm R, Waltimo O: Combined dexamethasone and low-molecular-weight dextran in acute brain infarction: Double-blind study. *Br Med J* 1976;2:1409-1410
10. Mulley G, Wilcox RG, Mitchell JRA: Dexamethasone in acute stroke. *Br Med J* 1978;2:994-996
11. Duncan PW, Propst M, Nelson SG: Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther* 1983;63:1606-1610
12. Sisk C, Ziegler DK, Zileli T: Discrepancies in recorded results from duplicate neurological history and examination in patients studied for prognosis in cerebrovascular disease. *Stroke* 1970;1:14-18
13. Tomesello F, Mariani F, Fieschi C, Argentino C, Bono G, DeZanche L, Inzitari D, Martini A, Perrone P, Sangiovanni G: Assessment of inter-observer differences in the Italian multicenter study on reversible cerebral ischemia. *Stroke* 1982;13:32-35
14. Shinar D, Gross CR, Mohr JP, Caplan LR, Price TR, Wolf PA, Hier DB, Kase LS, Fishman I, Wolf CL, Kunitz SC: Interobserver variability in the assessment of neurologic history and examination in the Stroke Data Bank. *Arch Neurol* 1985;42:557-565
15. Coté R, Hachinski VC, Shurvell BL, Norris JW, Wolfson C: The Canadian neurological scale: A preliminary study in acute stroke. *Stroke* 1986;17:731-737
16. Norris JW: Comment on "study design of stroke treatments" (letter). *Stroke* 1982;13:527-528
17. Oxbury JM, Greenhall RCD, Grainger KMR: Predicting the outcome of stroke: Acute cerebral infarction. *Br Med J* 1975;3:125-127
18. Sugiura K, Muranka K, Chishiki T, Baba M: The Edinburgh-2 Coma Scale: A new scale for assessing impaired consciousness. *Neurosurgery* 1983;12:411-415
19. The measurement of interrater agreement, in Fleiss JL: *Statistical Methods for Rates and Proportions*. New York, John Wiley & Sons, Inc, 1981
20. Hollander M, Wolfe DA: *Nonparametric Statistical Methods*. New York, John Wiley & Sons, 1973, pp 191-192
21. Brott T, Marler JR, Olinger CP, Adams HP Jr, Tomsick T, Barsan WG, Biller J, Eberle R, Hertzberg V, Walker M: Measurements of acute cerebral infarction: Lesion size by computed tomography. *Stroke* 1989;20:871-875
22. Spence JD, Donner A: Problems in design of stroke treatment trials. *Stroke* 1982;13:94-99
23. Feinstein AR: *Clinimetrics*. New Haven/London, Yale University Press, 1987
24. Jongbloed L: Prediction of function after stroke: A critical review. *Stroke* 1986;17:765-775
25. Goldstein LB, Bertels C, Davis JN: Interrater reliability of the NIH stroke scale. *Arch Neurol* 1989;46:660-662

KEY WORDS • cerebral infarction • clinical trials • neurologic examination