Examining Outcome Measures in a Clinical Study of Stroke

S.L. Wood-Dauphinee, PhD, PT, J. Ivan Williams, PhD, and S.H. Shapiro, PhD

We investigated the relation between outcome and sample size for six selected stroke outcome measures to assist investigators in selecting end points for stroke studies. Data from a clinical trial of 167 stroke patients assessed shortly after admission to the hospital and 5 weeks later provided information on clinical, motor, and functional outcomes measured using a neurologic status scale, a stroke severity scale, the Fugl-Meyer Scale, the Barthel Index, and the activities of daily living and cognition subscales of the Level of Rehabilitation Scale. Data were examined using Pearson correlation coefficients and power analyses. All measures were significantly correlated. There was also substantial congruency between the subscales of a measure and its total score. The measures had variable efficiencies; the Barthel Index was the most efficient and therefore required the fewest subjects to identify a significant effect. These data suggest that careful consideration must be given to the choice of stroke outcome measures in terms of their numbers, interrelationships, and statistical properties, as these factors have important implications for the design, analysis, and conduct of clinical stroke studies. (Stroke 1990;21:731-739)

In a 1986 editorial, Gresham noted several reasons for the growing interest in research on stroke outcome. First, when added to information on stroke incidence, data on stroke survival and disability provide a comprehensive profile of the impact of stroke that is important in planning health and social services. Second, information concerning those factors that influence survival or functional recovery after a stroke provides the framework within which death or the course of functional improvement of stroke patients can be predicted. And third, such research provides a baseline against which the results of controlled clinical trials, the most rigorous method of comparing the effects of various treatments, can be evaluated.

The usefulness of research on stroke outcome is easy to accept. The difficult task is selecting the clinical end points and the appropriate measures of outcome. Mortality, laboratory values, symptoms and signs of neurological dysfunction, and complications following the acute episode are the conventional assessments of outcome. More recently, importance has been given to evaluating the patient's functional performance at home and in the community in terms of motor, psychological, social, cognitive, and vocational performance and the effects of stroke on the quality of life of the patient and his or her family.

The dilemma of choosing outcome measures has increased as the end points have broadened. Assessing mortality or even the presence/absence of a neurologic sign is reasonably straightforward. Assessing the magnitude of a symptom, an aspect of functional performance, or the degree of disability is more difficult. The most complex assessments are related to the impact of stroke on daily functioning as this involves the physical, psychological, and social consequences of the disease and its treatment. Given the range and scope of stroke outcome variables and the variety of stroke outcome measures that are available, the researcher has a problem in choosing those to employ. One choice is to evaluate and report all clinically relevant outcomes. The argument is that because results of a treatment regimen may be favorable for some outcomes but adverse for others, all outcomes should be described. The use of multiple measures permits a study of the relations between them, and this can be particularly important for measures that have not been standardized.

If multiple measures are used, the investigator must have rules for deciding about the impact of the disease and the intervention. This decision can be...
arrived at by arranging the measures in a hierarchy of importance and weighting their results accordingly. An alternative approach is to combine all the information into an unweighted summary index score. However, the probability of a spurious conclusion increases with the number of measures. For example, with 20 measures and a significance level of \( \leq 0.05 \), one outcome is expected to be statistically significant by chance alone. Statistical procedures designed to correct for multiple measures increase the sample sizes required.

Sample sizes required for each measure also vary according to the statistical properties of the measure, particularly its reliability. In consequence, it is difficult to determine the precise number of subjects necessary to detect a significant effect. Finally, the psychometric properties of a measure (reliability, validity, and responsiveness) influence the results, adding to the complexity of interpretation.

The most widely advocated procedure is to select one end point as the primary focus of a study and to consider secondary outcomes that reflect milestones of clinical response or effects of the interventions. In the interest of conserving research resources and enhancing patient and staff compliance with the measurement regimes, the best advice may be "to choose as many as necessary and as few as possible." Factors influencing the choice of measures include the specific objective of the study, the availability of relevant standardized measures, the time and costs of administering them, and the willingness of staff and patients to use them. It is helpful to start with an understanding of the conceptual framework of the outcome variables and their relations. In selecting specific measures of the outcome variables of interest, one must consider the statistical properties of the measures and the implications for design of the study, analysis of the results, and sample size requirements. This information can be used to avoid redundancy and to select the most important outcomes as well as the most efficient measures for a specific study.

We had two related objectives. The first was to explore associations among outcome measures in a clinical study of stroke. The second objective was to examine the statistical properties of the measures and the resulting sample sizes required to detect a significant treatment effect. We also make some practical suggestions for investigators to consider when choosing measures.

**Subjects and Methods**

This paper is based on secondary analyses of data from a clinical study of stroke. Patients participating in the original investigation were enrolled in a two-phased, randomized controlled trial of interdisciplinary team care for hospitalized persons with acute stroke. The patients and study methods have been described in detail elsewhere. In brief, 172 patients with a motor or sensory deficit secondary to a thromboembolic or intracranial hemorrhagic vascular incident and no residual deficit from any previous cerebrovascular event were entered into the study. After obtaining baseline information on 42 stroke patients receiving conventional care in a general hospital, the remaining 130 stroke patients were stratified according to suspected prognosis and randomly assigned to either traditional or team care. Traditional care consisted of the care and services routinely available to patients in an acute-care hospital, ordered on an an ad hoc basis by the patient's personal physician and delivered by hospital personnel. Team care included individualized, comprehensive, and integrated programs of medical and nursing care; physical, occupational, and speech therapy; and social service delivered by stroke specialists who were members of the team. Assessments of the patients by trained evaluators permitted comparisons of survival, motor performance, and functional abilities between the traditional- and team-care groups. Patients were evaluated shortly after admission to the hospital and 5 weeks after the stroke.

For this paper, the three groups (conventional, traditional, and team care) were combined. The clinical, motor, and functional variables previously reported, as well as unused data from a stroke severity scale, a cognitive measure, and an activities of daily living (ADL) index were considered. These latter variables were measured at the same times and by the same evaluators as those previously reported.

We measured neurologic status of each patient on admission and at the 5-week follow-up. This measure divides the neurologic findings into five categories: mentation, motor nerves, cranial nerves, sensation, and reflexes. Patients are graded according to the most severe deficit in each category, and the grades are summed to provide an overall score. A patient with no neurologic deficit receives 44 points and a patient close to death 0. This neurologic status measure has predictive validity in terms of survival, is simple to use, and gives similar results when scored by different examiners.

We measured the severity of each patient's stroke on admission and at follow-up using a 10-point scale developed at McMaster University to judge the severity of a stroke. This measure incorporates neurologic findings of signs and symptoms with functional performance in terms of self-care ability, ambulation, and continence into a single index. A score of 0 means no symptoms, signs, or neurologic impairment and a score of 9 denotes death. A modified version of this 10-point scale was used in the EC/IC Bypass Study, in which the chance-corrected agreement among evaluators was 84%.

We assessed each patient's motor performance on admission and at follow-up using a measure quantified by Fugl-Meyer et al assuming that motor function improves after stroke in a predictable sequence. The Fugl-Meyer Scale includes five domains: upper extremity, lower extremity, balancing ability, sensation, and range of motion. A three-point ordinal scale (0, cannot perform; 1, performs partially; 2, performs fully) is applied to each item in a...
stroke, we living at home 5 weeks after their stroke, we community.34 has been used in a study of stroke patients living in the ADL and 0.88 for the cognition subscales. LORS a spouse give interinformant correlations of 0.82 for sons in the same professional discipline. A nurse and the neurologist scored each patient's neurologic status and stroke severity. After the patient's neurologic status had stabilized (3–5 days after admission), the admission assessments of the Fugl-Meyer Scale and the Barthel Index were made by one of two trained evaluators. The principal evaluator was a physical therapist with extensive experience in the assessment of neurological patients. Of 277 assessments conducted, she performed 271. In her absence, assessments were made by another experienced and trained physical therapist. Both evaluators were blinded to the hypotheses being tested and to the patient's group.

Within a few days after admission, information on sociodemographic characteristics; medical history including cerebrovascular disease and comorbidity; handedness; and side, site, and type of cerebral lesion was obtained from each patient's chart. Missing information was obtained through a brief interview with the patient or family. Income levels were estimated from 1975 census tract data according to address.

Five weeks after the stroke, the date and cause of death for those patients who died were obtained from the chart. The surviving patients were assessed on the Fugl-Meyer Scale, the Barthel Index, and the ADL and cognition subscales of LORS by an evaluator. One of the investigators (unblinded) also repeated the neurologic status and stroke severity assessments originally performed by a study neurologist. The follow-up assessment was conducted in the acute-care hospital, the patient's home, or a convalescent or rehabilitation institution.

Descriptive statistics of the patients' sociodemographic and clinical characteristics on admission were calculated. Pairwise relations between neurologic status, stroke severity, Fugl-Meyer Scale, and Barthel Index scores on admission were examined via scatter plots of the individual patients' scores and Pearson correlation coefficients. Similarly, the pairwise associations between neurologic status, stroke severity, Fugl-Meyer Scale, Barthel Index, and ADL and cognition LORS subscale scores at follow-up were investigated for the survivors using the same statistical approaches.

In addition, using both the mean change scores (admission to follow-up) and a hypothetical score change of 33.3% (one we deemed to be clinically relevant) we calculated the sample size required to detect a significant treatment effect for each using the coefficient of variation (CV) and effect size.35

As variation within a group increases, differences between the means of groups become more difficult to detect. The efficiency of a measure can be estimated by its CV, the group standard deviation (SD) divided by the respective group mean. For example, data with an SD of 10 and a mean of 100 has a CV of 0.10. A low CV indicates an efficient measure. As group SDs approach or surpass their means, that is, as group CVs near 1, the measure becomes inefficient. Another method of judging differences between groups is effect size, the difference between...
the group means divided by the SD. Effect size can be viewed as the amount of change expressed in SD units and thus provides a standardized assessment of the difference between groups. An effect size of 0 supports the null hypothesis of no difference. As effect size increases, the data become more suggestive of a difference between two groups. An effect size of 0.10 is small; one of ≥0.70 is large. When a new treatment is compared with a traditional one, as is often the case in rehabilitation, the effect size is generally moderate.36 Differences between groups are assessed statistically in relation to the standard error of the mean (SEM), the group SDs divided by the square root of the sample size (n). For a given difference between groups, the larger the SDs the larger the n's must be for the difference to be significant. In calculating n, we specify the minimum effect size, set the significance and power levels, and then estimate the number of subjects required to detect a difference. The sample size calculations are based on a one-tailed t test with an α level of 0.05 and a power of 0.80, but do not take into account patients who die or are lost to follow-up. The number of such patients would have to be estimated by the investigator and would increase the sample size required.

Results

Of the 172 patients, 167 remained in the study for 5 weeks or until death; 48 died. Table 1 presents the sociodemographic characteristics of the 167 patients who completed the study. Their mean±SD age was 73.7±10.2 years. Table 2 displays the clinical characteristics of the 167 patients. Most had one or more of the serious health problems frequently found in elderly individuals (data not shown).15

Table 3 shows the mean clinical (neurologic status and stroke severity), motor (Fugl-Meyer upper extremity, lower extremity, balancing ability, and total), and functional (Barthel self-care ability, mobility, and total) scores at admission and follow-up as well as the maximum scores possible for each measure. As expected, the survivors performed better than the entire group on all measures at admission. Scores had improved markedly by the 5-week follow-up.

As seen in Table 4, the scores at admission of all measures are moderately to highly and significantly correlated. As anticipated, scores on the subscales of both the Fugl-Meyer Scale and the Barthel Index are highly correlated with total scores on their parent measures. Coefficients for both subscale and total scores are provided as both are used as outcomes clinically and in research. Negative coefficients for stroke severity are due to the fact that lower scores indicate less impairment. The stroke severity score, which combines both neurologic and functional assessments, was highly correlated with the neurologic status score but to a lesser degree with scores on the motor or functional measures. Scores on the balancing ability subscale of the Fugl-Meyer Scale correlated with those on the motor and functional measures but to a lesser extent with those on the two clinical measures.

Table 5 provides correlation coefficients for the follow-up data and includes two additional variables, the ADL and cognition subscales of LORS, as these measures were made only at follow-up. In general, the correlations at follow-up are similar to but stronger than those at admission.

Table 6 displays the correlation coefficients for the mean change scores. The correlations are all still significant, but they are weaker than those at admission or at follow-up.

In addition to the correlation coefficients, we also reviewed the scatter plots (data not shown). The plots gave no reason to suspect nonlinearity or that nonzero correlations arose artificially.

Table 7 lists the mean±SD change scores and the corresponding CVs, effect sizes, and sample sizes required to detect a treatment effect of 33Vs% for

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50.3</td>
</tr>
<tr>
<td>Female</td>
<td>49.7</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>50.3</td>
</tr>
<tr>
<td>Divorced/separated</td>
<td>2.4</td>
</tr>
<tr>
<td>Widowed</td>
<td>43.7</td>
</tr>
<tr>
<td>Never married</td>
<td>3.6</td>
</tr>
<tr>
<td>Primary language spoken</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>53.9</td>
</tr>
<tr>
<td>French</td>
<td>14.9</td>
</tr>
<tr>
<td>Yiddish</td>
<td>12.0</td>
</tr>
<tr>
<td>Other</td>
<td>19.2</td>
</tr>
<tr>
<td>Usual occupation</td>
<td></td>
</tr>
<tr>
<td>Professional/managerial</td>
<td>12.6</td>
</tr>
<tr>
<td>Clerical/service/sales</td>
<td>15.6</td>
</tr>
<tr>
<td>Crafts/production</td>
<td>13.1</td>
</tr>
<tr>
<td>Laborer</td>
<td>8.4</td>
</tr>
<tr>
<td>None stated</td>
<td>43.1</td>
</tr>
<tr>
<td>Not determined</td>
<td>7.2</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>97.6</td>
</tr>
<tr>
<td>Other</td>
<td>2.4</td>
</tr>
<tr>
<td>Living arrangement</td>
<td></td>
</tr>
<tr>
<td>Family</td>
<td>69.5</td>
</tr>
<tr>
<td>Alone</td>
<td>22.7</td>
</tr>
<tr>
<td>Others</td>
<td>12.2</td>
</tr>
<tr>
<td>Institution</td>
<td>6.6</td>
</tr>
<tr>
<td>Family income (1979–80)</td>
<td></td>
</tr>
<tr>
<td>&lt;$5,000</td>
<td>1.8</td>
</tr>
<tr>
<td>$5,000–$10,000</td>
<td>57.5</td>
</tr>
<tr>
<td>$10,000–$15,000</td>
<td>29.9</td>
</tr>
<tr>
<td>&gt;$15,000</td>
<td>10.8</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>4.2</td>
</tr>
<tr>
<td>Employed full time</td>
<td>11.3</td>
</tr>
<tr>
<td>Employed part time</td>
<td>2.4</td>
</tr>
<tr>
<td>Homemaker</td>
<td>12.0</td>
</tr>
<tr>
<td>Unable to work</td>
<td>0.6</td>
</tr>
<tr>
<td>Retired</td>
<td>69.5</td>
</tr>
</tbody>
</table>

Downloaded from http://stroke.ahajournals.org/ by guest on June 24, 2017
TABLE 2. Clinical Characteristics of 167 Stroke Patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of cerebrovascular disease</td>
<td></td>
</tr>
<tr>
<td>Nil</td>
<td>63.5</td>
</tr>
<tr>
<td>Transient ischemic attacks</td>
<td>20.9</td>
</tr>
<tr>
<td>Stroke (no residual)</td>
<td>15.6</td>
</tr>
<tr>
<td>Location of lesion</td>
<td></td>
</tr>
<tr>
<td>Hemisphere</td>
<td>94.0</td>
</tr>
<tr>
<td>Brainstem</td>
<td>5.4</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>0.6</td>
</tr>
<tr>
<td>Level of consciousness</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>39.5</td>
</tr>
<tr>
<td>Somnolent</td>
<td>26.3</td>
</tr>
<tr>
<td>Stuporous</td>
<td>18.6</td>
</tr>
<tr>
<td>Comatose</td>
<td>15.6</td>
</tr>
<tr>
<td>Side of weakness</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>49.7</td>
</tr>
<tr>
<td>Left</td>
<td>48.5</td>
</tr>
<tr>
<td>Both</td>
<td>1.8</td>
</tr>
<tr>
<td>Handedness</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>89.2</td>
</tr>
<tr>
<td>Left</td>
<td>6.6</td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>0.6</td>
</tr>
<tr>
<td>Not determined</td>
<td>3.6</td>
</tr>
<tr>
<td>Final diagnosis</td>
<td></td>
</tr>
<tr>
<td>Thrombus</td>
<td>56.9</td>
</tr>
<tr>
<td>Embolus</td>
<td>20.3</td>
</tr>
<tr>
<td>Thromboemboli</td>
<td>16.8</td>
</tr>
<tr>
<td>Hemorrhage</td>
<td>6.0</td>
</tr>
</tbody>
</table>

each measure. The CVs range from 0.80 to 2.67, suggesting that the measures were not reliable, the study population was heterogeneous, or both. The literature demonstrates that most of these measures are reliable, so the principal reason for their relative inefficiencies was population heterogeneity. In addition, the subjects were more heterogeneous on some measures (e.g., neurologic status) than on others (e.g., total score on the Barthel Index).

We assumed that if treatment increased the score change by one third of that in the control group, the finding would be clinically significant. In Table 7 such a one third improvement in score change was used to calculate effect size, which was small for neurologic status and moderate for the Barthel Index. The sample sizes required for these effect sizes to be deemed statistically significant are also shown in Table 7. If the total score on the Barthel Index were the primary measure of outcome in a classical two-group randomized design, the control and treatment groups would each require 72 subjects; 265 subjects would be required for each group using neurologic status score as the measure.

Discussion

Patients enrolled in the study were elderly individuals admitted to the medical wards of a general hospital for the management of acute stroke. Data from a randomized treatment trial provided the opportunity to explore the relations between different measures of clinical outcome. As expected, the correlations between measures of neurologic status, stroke severity, motor performance, cognition, and functional capacity were all significant. The correlation coefficients provide important evidence concerning the relations between motor, functional, and balancing ability.

For both the Fugl-Meyer Scale and the Barthel Index, the subscale scores and total scores were highly correlated at both the admission and follow-up assessments. Each subscale assesses the impact of stroke on a particular dimension of functioning, and when combined the total score indicates the systemic

TABLE 3. Mean±SD Clinical, Motor, and Functional Scores of Stroke Patients at Admission and at Follow-up

<table>
<thead>
<tr>
<th>Measure</th>
<th>Maximum possible</th>
<th>All patients (n=167 or 158*)</th>
<th>Survivors (n=119)</th>
<th>At follow-up (n=119 or 118t)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurologic status16</td>
<td>44‡</td>
<td>24.49±9.5</td>
<td>29.82±7.78</td>
<td>32.38±9.1†</td>
</tr>
<tr>
<td>Stroke severity18</td>
<td>9</td>
<td>5.86±1.5</td>
<td>5.38±1.4</td>
<td>4.52±1.9†</td>
</tr>
<tr>
<td><strong>Motor (Fugl-Meyer Scale19)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper extremity</td>
<td>66</td>
<td>24.21±23.7*</td>
<td>30.45±23.6</td>
<td>41.32±25.7†</td>
</tr>
<tr>
<td>Lower extremity</td>
<td>34</td>
<td>13.65±11.2*</td>
<td>19.07±10.4</td>
<td>23.29±10.9†</td>
</tr>
<tr>
<td>Balancing ability</td>
<td>14</td>
<td>3.96±3.9*</td>
<td>5.10±3.7</td>
<td>8.28±4.3†</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>43.54±36.3*</td>
<td>54.32±34.7</td>
<td>72.82±39.6†</td>
</tr>
<tr>
<td><strong>Functional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barthel Index27</td>
<td>53</td>
<td>16.65±18.2*</td>
<td>21.46±18.3</td>
<td>35.39±19.6</td>
</tr>
<tr>
<td>Mobility</td>
<td>47</td>
<td>6.83±10.7*</td>
<td>9.01±11.5</td>
<td>25.53±18.7</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>23.48±27.1*</td>
<td>30.47±27.7</td>
<td>60.92±37.1</td>
</tr>
<tr>
<td>Level of Rehabilitation Scale29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities of daily living</td>
<td>100</td>
<td>...</td>
<td>...</td>
<td>56.18±37.3</td>
</tr>
<tr>
<td>Cognition</td>
<td>100</td>
<td>...</td>
<td>...</td>
<td>67.28±37.0</td>
</tr>
</tbody>
</table>

*Nine patients died before motor and functional evaluations were made.
†Incomplete follow-up in one survivor.
‡Score of 0 indicates no impairment.
TABLE 4. Pearson Correlation Coefficients for Clinical, Motor, and Functional Scores of Stroke Patients at Admission

<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Neurologic status(^{16})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Stroke severity(^{18})</td>
<td>-0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motor (Fugl-Meyer Scale(^{19}))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Upper extremity</td>
<td>0.65</td>
<td>-0.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Lower extremity</td>
<td>0.68</td>
<td>-0.64</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Balancing ability</td>
<td>0.69</td>
<td>-0.70</td>
<td>0.81</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Total</td>
<td>0.70</td>
<td>-0.66</td>
<td>0.97</td>
<td>0.90</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Functional (Barthel Index(^{27}))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Self-care ability</td>
<td>0.69</td>
<td>-0.74</td>
<td>0.73</td>
<td>0.74</td>
<td>0.79</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Mobility</td>
<td>0.55</td>
<td>-0.61</td>
<td>0.67</td>
<td>0.69</td>
<td>0.83</td>
<td>0.73</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>I. Total</td>
<td>0.68</td>
<td>-0.73</td>
<td>0.75</td>
<td>0.77</td>
<td>0.86</td>
<td>0.82</td>
<td>0.97</td>
<td>0.90</td>
</tr>
</tbody>
</table>

\(^{n}=167\) for A and B, 158 for C-I (nine patients died before evaluation of motor and functional measures). \(p<0.01\) for all correlations; mean \(|r|\)=0.754, SD of \(r\)=0.10.

TABLE 5. Pearson Correlation Coefficients for Clinical, Motor, and Functional Scores of Surviving Stroke Patients at 5-Week Follow-up

<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Neurologic status(^{16})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Stroke severity(^{18})</td>
<td>-0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motor (Fugl-Meyer Scale(^{19}))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Upper extremity</td>
<td>0.85</td>
<td>-0.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Lower extremity</td>
<td>0.86</td>
<td>-0.79</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Balancing ability</td>
<td>0.83</td>
<td>-0.82</td>
<td>0.86</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Total</td>
<td>0.88</td>
<td>-0.80</td>
<td>0.98</td>
<td>0.95</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Functional</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barthel Index(^{27})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Self-care ability</td>
<td>0.84</td>
<td>-0.84</td>
<td>0.79</td>
<td>0.86</td>
<td>0.88</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Mobility</td>
<td>0.80</td>
<td>-0.83</td>
<td>0.80</td>
<td>0.87</td>
<td>0.92</td>
<td>0.86</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. Total</td>
<td>0.85</td>
<td>-0.86</td>
<td>0.82</td>
<td>0.89</td>
<td>0.93</td>
<td>0.88</td>
<td>0.97</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td><strong>Level of Rehabilitation Scale(^{33})</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Activities of daily living</td>
<td>0.83</td>
<td>-0.87</td>
<td>0.80</td>
<td>0.86</td>
<td>0.92</td>
<td>0.86</td>
<td>0.94</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>K. Cognition</td>
<td>0.82</td>
<td>-0.82</td>
<td>0.68</td>
<td>0.75</td>
<td>0.74</td>
<td>0.73</td>
<td>0.85</td>
<td>0.77</td>
<td>0.84</td>
</tr>
</tbody>
</table>

\(^{n}=119\) for G–K, 118 for A–F (incomplete follow-up in one patient). \(p<0.01\) for all correlations; mean \(|r|\)=0.854, SD of \(r\)=0.064.

TABLE 6. Pearson Correlation Coefficients of Mean Changes for Clinical, Motor, and Functional Scores of Stroke Patients Surviving 5 Weeks

<table>
<thead>
<tr>
<th>Measure</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Neurologic status(^{16})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Stroke severity(^{18})</td>
<td>-0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Motor (Fugl-Meyer Scale(^{19}))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Upper extremity</td>
<td>0.45</td>
<td>-0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Lower extremity</td>
<td>0.58</td>
<td>-0.42</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Balancing ability</td>
<td>0.59</td>
<td>-0.57</td>
<td>0.56</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Total</td>
<td>0.58</td>
<td>-0.46</td>
<td>0.94</td>
<td>0.75</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Functional (Barthel Index(^{27}))</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Self-care ability</td>
<td>0.51</td>
<td>-0.42</td>
<td>0.51</td>
<td>0.57</td>
<td>0.52</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. Mobility</td>
<td>0.46</td>
<td>-0.55</td>
<td>0.46</td>
<td>0.41</td>
<td>0.65</td>
<td>0.55</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>I. Total</td>
<td>0.56</td>
<td>-0.57</td>
<td>0.57</td>
<td>0.57</td>
<td>0.68</td>
<td>0.68</td>
<td>0.85</td>
<td>0.87</td>
</tr>
</tbody>
</table>

\(^{n}=119\) for G–I, 118 for A–F (incomplete follow-up in one patient). \(p<0.01\) for all correlations; mean \(|r|\)=0.578, SD of \(r\)=0.130.
effects of the disease. This means that investigators can obtain useful information using one or two subscales rather than the complete measures when patients are limited in their ability to respond or when there are constraints on resources.

Similarly, the stroke severity measure, which assesses impairment in ambulation and self-care ability, correlates more strongly with the Barthel Index and ADL subscale of LORS than with the measures of motor performance. The ADL subscale correlates best with the Barthel Index. These outcome measures are related in that they show the effects of stroke on particular dimensions while indicating the overall impact of the stroke on the individual.

The measure of balancing ability correlates highly with measures of both motor performance and ADL. Our correlations of upper and lower extremity performance with balancing ability are greater than those reported by Fugl-Meyer.37 Badke and Duncan38 also determined that the lower extremity subscale of the Fugl-Meyer Scale indicated proportional electromyographic activity when hemiplegic patients were asked to balance on a moving platform. These authors concluded that the lower the score, the greater the abnormality in the organization of postural responses. Other investigators19,40 have found weaker associations between balancing ability and lower extremity than upper extremity motor function. These authors failed to explain their negative findings, but they noted that good standing balance is required for other purposeful motor activities. Finally, several researchers have noted that the ability to maintain an upright position is strongly related to ADL skills.19,40,41

The relation between balancing ability and cognition is also moderately strong. Although this relation has not been studied extensively in stroke patients, Tinetti et al42 found that cognitively impaired individuals are at greater risk of falling than normal elderly individuals, and balance is related to falls.43 Scores on the cognition subscale of LORS are more strongly related to functional measures than to measures of motor performance. Three studies44–46 have noted a relation between visual-spatial performance and ADL skills. Associations between constructive apraxias and body scheme disorders and functional abilities47 as well as between auditory attention and overall ADL functioning have also been reported.45

As a measure of ADL performance may reflect psychological and social adjustment in addition to physical ability, many stroke studies include both types of outcome measures. We found a strong relation between motor ability and functional performance, and this confirms the findings of a number of studies.41,48–51

This raises the question of whether measures of both motor performance and functional ability need to be employed in a study. Any trained individual can reliably assess a patient on the Barthel Index, while a trained therapist is required to administer the Fugl-Meyer instrument.40 As seen in Table 7, the CVs for the Fugl-Meyer Scale and its subscales are >1.0 while the CVs for the Barthel Index and its subscales are <1.0. The Fugl-Meyer Scale is a less efficient measure of clinical change, hence it requires more subjects than the Barthel Index. This suggests that priority should be given to the Barthel Index unless the Fugl-Meyer Scale is needed to test particular interventions.

Correlations between the measures at follow-up were higher than those at admission, and there are several possible explanations for this finding. At follow-up, experience may have improved the performance of the evaluators and the patients may have been more cooperative. At admission, severely affected patients had very low scores, thereby increasing the variability and suppressing associations. Other patients with low scores at admission may have, through training, learned to optimize their functional activities even if there had been little change in their capacity for motor performance. All of these factors may have contributed to the better correlations at follow-up.

Correlations for the mean score changes, on the other hand, were lower than for either evaluation time. Because the score changes incorporate the

<table>
<thead>
<tr>
<th>Measure</th>
<th>Score change (mean±SD)</th>
<th>Coefficient of variation</th>
<th>Effect size</th>
<th>Sample size (per group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neurologic status16</td>
<td>2.62±7.00</td>
<td>2.67</td>
<td>0.12</td>
<td>810</td>
</tr>
<tr>
<td>Stroke severity18</td>
<td>−0.87±1.34</td>
<td>1.54</td>
<td>0.22</td>
<td>265</td>
</tr>
<tr>
<td>Motor (Fugl-Meyer Scale27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper extremity</td>
<td>11.09±18.49</td>
<td>1.67</td>
<td>0.20</td>
<td>312</td>
</tr>
<tr>
<td>Lower extremity</td>
<td>4.28±7.68</td>
<td>1.79</td>
<td>0.19</td>
<td>358</td>
</tr>
<tr>
<td>Balancing ability</td>
<td>3.21±3.27</td>
<td>1.02</td>
<td>0.33</td>
<td>117</td>
</tr>
<tr>
<td>Total</td>
<td>18.73±25.55</td>
<td>1.36</td>
<td>0.24</td>
<td>208</td>
</tr>
<tr>
<td>Functional (Barthel Index27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-care ability</td>
<td>14.03±13.59</td>
<td>0.97</td>
<td>0.34</td>
<td>106</td>
</tr>
<tr>
<td>Mobility</td>
<td>16.53±14.75</td>
<td>0.89</td>
<td>0.37</td>
<td>90</td>
</tr>
<tr>
<td>Total</td>
<td>30.55±24.31</td>
<td>0.80</td>
<td>0.42</td>
<td>72</td>
</tr>
</tbody>
</table>
variability of both evaluation times, the strength of the associations may have diminished.

Choosing the appropriate measures of outcomes for a clinical study of stroke patients is an exercise in realism. There is a difference between what the investigator would like to include and that which is manageable. While the initial concern might be to miss nothing, crisp thinking is required to consider the implications of the measurement protocol on sample size, statistical analysis, costs in terms of both time and money, as well as encumbrance on the patient.

Measures must be related to the objectives of the study and must be suitable for assessing the effect of the intervention under investigation. The reliability, validity, and responsiveness of the measure to important clinical differences determine the quality of the data and the number of subjects required. The time, costs, and burden imposed on the patient and caregiver by the methods must be taken into account. Redundancy in measures increases time, costs, and burden and reduces the marginal utility of the information added by the redundant measure. The inclusion of a given measure should be justified by its proponent in terms of the relative difficulties in obtaining the information, the statistical properties of the measure, and the uses that can be made of the results.

References

KEY WORDS • cerebrovascular disorders • clinical trials • stroke outcome
Examining outcome measures in a clinical study of stroke.
S L Wood-Dauphinee, J I Williams and S H Shapiro

*Stroke*. 1990;21:731-739
doi: 10.1161/01.STR.21.5.731

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1990 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/21/5/731

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/