Outcomes studies in cerebrovascular neurosurgery

CARLOS E. SANCHEZ, M.D., CHRISTOPHER S. OGILVY, M.D.,
AND BOB S. CARTER, M.D., PH.D.

Cerebrovascular Surgery Unit, Neurosurgical Service, Massachusetts General Hospital

Successfully measuring cerebrovascular neurosurgery outcomes requires an appreciation of the current state-of-the-art epidemiological instruments, their specific relevance to surgical treatments and the underlying pathological entity, and ultimately the right set of questions for the next generation of studies. In this paper the authors address these questions with specific attention to measurement targets, individual modeling scales, and types of studies, all within a conceptual framework for specific disease models in their current state of outcomes modeling in cerebrovascular neurosurgery.

KEY WORDS • cerebrovascular disease • infarction • hemorrhage • aneurysm • cavernous malformation • outcome assessment

Abbriviations used in this paper: ADL = activity of daily living; AVM = arteriovenous malformation; CEA = carotid endarterectomy; GCS = Glasgow Coma Scale; GOS = Glasgow Outcome Scale; ICH = intracerebral hemorrhage; MMSE = Mini-Mental State Examination; NIH = National Institutes of Health; QOL = Quality of Life; RNL = Reintegration to Normal Living; SAH = subarachnoid hemorrhage; SF-36 = Short Form-36; SIP = Sickness Impact Profile.
features of study design contribute significantly to the usefulness of any outcomes analysis: 1) the methodological rigor of the study, proceeding along the continuum from unsystematic clinical observations, to observational studies with appropriate control groups, to the use of bias-reducing techniques such as randomization and prospective data collection; and 2) the cofactors studied to permit adequate subgroup analysis and therefore better prediction for groups of individuals with a given set of cofactors.

Each of these areas (methodological rigor, cofactors studied, and outcome measures used) contributes to the ultimate goal of understanding how a patient will fare over time. For example, in a population of patients with unruptured aneurysms, one would certainly like to know about differences in the posttreatment GOS scores according to the patient-specific factors of age, size, and location of the aneurysm. Better still would be to know the anticipated neuropsychological and QOL outcomes for coil placement compared with surgery according to these same patient-specific variables. Of course, improvement in any one of the aforementioned areas that contributes to a study’s value will increase the cost, time, and overall ability to perform the study. From the patient’s perspective, the goal of this progression is to move closer and closer to answering the question “How am I likely to do . . . specifically?”

In this paper we describe the state of the art of clinical cerebrovascular outcomes analyses and future directions for several disease entities of interest to cerebrovascular surgeons. Initial consideration is given to a review of the relevant outcomes and suitable instruments for cerebrovascular outcomes studies. Furthermore, there is an examination of the current level of application of these instruments in cerebrovascular disorders. Finally, in the cerebrovascular outcomes literature, there are many clinical series in which broad, often author-defined groupings of patient outcomes are examined. We hope that this review will shine a light on future directions for cerebrovascular surgery outcomes analyses.

WHAT SHOULD BE MEASURED IN CEREBROVASCULAR SURGERY OUTCOMES?

The question of what should be measured in cerebrovascular outcomes is not trivial. Certainly, there are some outcomes such as death that should be measured in nearly every condition treated by the cerebrovascular surgeon. On the other hand, some disease conditions will affect particular dimensions of a patient’s health, rendering only certain outcome measures important. The cognitive impairment of the patient with SAH is less relevant to the patient with a spinal cavernous malformation who is focused on mobility. Furthermore, within a given disease process there may be certain subgroups of patients for whom differing outcome measures may be appropriate. Patrick and Deyo have highlighted two different approaches to the assessment of health status; a generic and a disease-specific model. Under this model it would be appropriate to measure both a general outcome (for example, allowing a general assessment of QOL that can be used to compare a population with one disease with another) as well a disease-specific set of outcomes that allow for the standardized assessment of the unique set of impairments that are associated with cerebrovascular disorders.

McDowell and Newell have provided a comprehensive review of a variety of health outcome measures, including many measures with specific relevance to neurological disorders. Wade has reviewed a large number of the instruments available for outcomes analysis in neurological disorders, including cerebrovascular disease. In the following sections we discuss a number of measurement areas and instruments available for both general and disease-specific assessment of outcome in cerebrovascular surgery. The instruments reviewed are summarized and referenced in Table 1 in terms of assessment area, measures of reliability, and methods of validation.

MORTALITY DATA

Although seemingly one of the most basic of clinical outcome assessments, mortality data are not always included in reports of cerebrovascular outcomes; a recent review showed that mortality was assessed in only 66% of 55 acute stroke trials, despite the importance and robustness of this measure. This outcome measure has relevance to nearly every disease process of interest to the cerebrovascular surgeon, including cerebral infarction, ICH, and SAH. This measure is useful in large-scale epidemiological studies as well as in smaller investigational studies of specific therapeutic modalities or protocols.

The primary problems in interpreting mortality data include: 1) loss of patients to follow up or incomplete ascertainment, primarily in smaller retrospective studies with limited resources for follow up; 2) insufficient follow-up duration, especially in regard to late deaths that may be relevant when examining the effects of disease interventions that are expected to have a long-term effect over a period of years or have a delayed onset of action (for example, radiosurgery for AVM); and 3) errors in death assignment.

Large-scale epidemiological studies have been published reporting population-based case-fatality rates for the three major subtypes of stroke, including SAH, ICH, and subtypes of cerebral infarction. These types of studies provide important natural history data across populations. However, their breadth may preclude in-depth assessment of cofactors. Prospective case ascertainment in tertiary care centers or hospital-based surveillance suffers from potential selection bias but may be useful for analysis of factors associated with mortality data within specific subgroups of patients because of the ability to collect in-depth information on these patients.

A common strategy for death reporting includes the number of deaths at 28 days after an ictus or intervention. This is not adequate for those conditions or treatments wherein a late change in the patient’s status is possible or likely. Reporting at late time intervals may require additional surveillance measures to determine whether death has occurred. In the US, one information resource is the Social Security Administration, which maintains a Death Master File of approximately 76 million persons who have had termination of Social Security benefits. The Death Master File can be purchased directly from the US Department of Commerce (for details go to www.ssa.gov).
Outcomes studies in cerebrovascular neurosurgery

Once a death has been identified, it can be difficult to assign its cause accurately, especially if this does not occur in the hospital. The positive predictive value of death certificate assignment of stroke as a cause of death has been estimated to be between 60 and 100%, with sensitivities ranging from 63.7% in the Framingham Heart Study\textsuperscript{6,118} to 70% in the Minnesota Heart Survey.\textsuperscript{63} Despite the aforementioned difficulties with death as an outcome measure, it remains one of the most reproducible and valid measures of outcome across populations.

Morbidity Assessment: A Conceptual Framework—Impairments, Activities, and Participation

The International Classification of Functioning and Disability, also known as the International Classification of Functioning, Disability, and Health (or ICF), provides a three-tiered framework for understanding morbidity assessment for cerebrovascular outcomes. In the classification provided by the aforementioned instrument, the morbidity associated with a particular pathological entity may be described in terms of three levels of outcome, including the following: 1) impairments (symptoms and signs of illness); 2) restrictions in activities (functions and observed behaviors); and 3) restrictions in participation (social roles or positions).\textsuperscript{60} Thus, a patient recovering from an SAH may have a residual hemiparesis (impairment), which precludes lifting (activity/function), which in turn prevents return to work (restriction in participation). It has been suggested that the most useful outcome measures relate to a single one of these three levels of outcome because of the difficulty in interpreting results from measures that reflect more than one level.

More advanced cerebrovascular outcomes studies will include not only neurologically specific impairment (for example, the NIH Stroke Scale) but also activity restriction (for example, the Barthel Index) and general QOL/participation-based outcome assessment as part of an overall framework. Measures that occur at the level of activities are the most important primary outcome measure. Impairment measures should be included in any study because of their importance in terms of stratifying case mix and as surrogate outcome measures, but the primary focus of the patient remains on activity. Third-tier outcomes, at the level of participation/QOL, are subject to a variety of extraneous factors, including personal background and physical and social environment. These influencing factors render QOL measures less appropriate for specific comparative assessments of different therapies. Nonetheless, these measures are attractive conceptually because of their ability to describe the effect of a disease and associated interventions on the overall health of a patient. Roberts and Counsell\textsuperscript{69} reviewed the use of measures of impairment, activity restriction, and participation/QOL in an analysis of 174 acute stroke trials for the Cochrane Stroke Group’s database from 1966 to 1995. Impairment was measured in 76% of trials, whereas activity restriction was measured in only 42% of studies and QOL in just 2%.

Assessment of Impairment-Level Outcomes in Cerebrovascular Surgery

In studies of patients with ischemic stroke, SAH, and other disorders, there are important associations between QOL/participation measures and the underlying levels of outcome, including impairment and activity limitation. Because of this, the reporting of outcomes at the level of impairments is important in any comprehensive outcomes survey. Stratification of activity and participation outcomes by impairments allows for improved comparison between populations. Three major areas of impairment in cerebrovascular disease include depressive symptoms, physical impairment, and cognitive dysfunction. The complex interrelationships between these impairments and activity- and participation-level outcomes have begun to be explored in patients who have suffered stroke and SAH. Several scales appropriate for assessment of impairment-level outcomes in cerebrovascular disorders are described in the following sections.

Stroke Scales. Cerebral infarction is the underlying pathological entity for much of the morbidity observed in cerebrovascular surgery. There is a considerable literature on measurement of the initial severity of cerebral infarction in terms of neurological impairment, and a variety of specialized scales have been developed for this purpose. Although no stroke scale fulfills all of the requirements for rigorous psychometric design, a few have been generally accepted because of their simplicity and content validity. The NIH Stroke Scale\textsuperscript{55} is a valid, reliable, and quickly administered test (5–8 minutes to complete) that is widely used in stroke research. Other recommended standardized scales include the Hemispheric Stroke Scale\textsuperscript{17} and the Canadian Neurological Scale,\textsuperscript{66} which are similarly easy to administer and provide some categorization of neurological impairment.

These scales are not been developed for use within patients who have suffered ischemic cerebral infarction. They have not been validated for use in other disorders such as SAH.

Cognitive Outcomes Instruments. Outcomes instruments that are used to assess cognition represent a specialized form of neurological impairment assessment. The MMSE\textsuperscript{66} is one of the most widely used quick screening measures for cognitive impairment. Normative values by population and age have been established and can be used as a reference in the absence of a specific control population.\textsuperscript{85} An alternative interview-based measure is the Telephone Interview for Cognitive Status, which is a valid alternative to the MMSE.\textsuperscript{56,64} The Blessed Dementia rating scales\textsuperscript{20} were first published in 1968 and also represent a quick screening tool for cognitive dysfunction with good validity testing against neuropathological findings. The GCS\textsuperscript{176} is the best known, most widely used scale to assess level of consciousness. It is simple, reliable, and valid and represents a valuable tool for general screening of the level of consciousness. For cerebrovascular disorders, it is useful for stratification of patients for analysis of activity-level outcomes.

The Wechsler Adult Intelligence Scale,\textsuperscript{100} which is the most widely used test of intelligence, has the following subscales: Verbal (information, comprehension, arithmetic, similarities, digit span, and vocabulary), and Performance (digit-symbol, picture completion, block design, picture arrangement, and object assembly). A selection of the subtests may be used, with prorated results. The

Neurosurg. Focus / Volume 22 / March, 2007 3
TABLE I

<table>
<thead>
<tr>
<th>Scale Category†</th>
<th>Quantitative Measure of Reliability‡</th>
<th>Method of Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barthel Index</td>
<td>activity test–retest reliability = 0.98</td>
<td>correlation coefficients b/w total ADL score &amp; Motricity Index arm, leg, &amp; total scores (data relating to patients seen at 3 wks): r = 0.729 (arm), 0.751 (leg), &amp; 0.774 (total); all p &lt; 0.001</td>
</tr>
<tr>
<td>Beck Depression Inventory</td>
<td>impairment κ = 0.88</td>
<td>predictive validity using crude odds ratio: 0.73 (death w/in 6 mos), 0.70 (any cardio- or cerebrovascular events w/in 6 mos), 1.60 (independence at 6 mos); p ≤ 0.001 for all</td>
</tr>
<tr>
<td>Beck Questionnaire</td>
<td>impairment valid, reliable, &amp; sensitive in physically healthy, psychiatrically ill patients</td>
<td>interrater agreement K w = 0.75 agreement btwn instrument &amp; criterion dx: κ = 0.93, 95% CI = 0.88–0.98</td>
</tr>
<tr>
<td>Blessed Dementia Rating Scale</td>
<td>impairment validity as diagnostic measure of depression w/ physical disability</td>
<td>correlates significantly w/ total nursing contact time in patients w/ traumatic brain &amp; spinal cord injury (r = 0.40–0.60)</td>
</tr>
<tr>
<td>Neuro-psychological Automated Battery</td>
<td>impairment test–retest reliability 0.75–0.80 only seen w/ match to sample-visual search task, accuracy measure for two-choice recognition task, &amp; avg no. of trials to success on the paired associate memory task</td>
<td>excellent discrimination (0.833, 95% CI = 0.82–0.846 under ROC curve) but poor calibration (Pearson’s χ² = 122 on 11 df)</td>
</tr>
<tr>
<td>Canadian Neurological Scale</td>
<td>impairment internal consistency using Cronbach α = 0.792</td>
<td>concurrent validity Pearson coefficient r = 0.89</td>
</tr>
<tr>
<td>Epilepsy Surgery Inventory–55</td>
<td>Cronbach α range = 0.76–0.88 w/ exception of social function (α = 0.68)</td>
<td>observed vs predicted classification b/w back-to-work &amp; still unemployed groups avg 76.6% correct</td>
</tr>
<tr>
<td>Functional Activities Questionnaire</td>
<td>activity interrater reliability (mean κ = 0.69), test–retest reliability w/ reported interclass correlations for both &gt; 0.90</td>
<td>performance strongly related (p &lt; 0.001) to 2 other independent measures of patient functioning</td>
</tr>
<tr>
<td>FIM</td>
<td>activity Cronbach α &gt; 0.90; excellent interrater &amp; test–retest reliabilities w/ reported interclass correlations for both &gt; 0.90</td>
<td>p values for construct validity describing treatment effects in following subscales: patient (0.017); caregivers (0.035); happiness (0.003); &amp; mastery (0.003)</td>
</tr>
<tr>
<td>GCS</td>
<td>activity moderate degrees of interrater agreement w/ GCS 32% for exact total GCS btwn ED MDs interobserver reliability r = 0.95</td>
<td>correlations ranged from −0.66 to −0.93 on variety of cognitive screening tests</td>
</tr>
<tr>
<td>Hemispheric Stroke Scale</td>
<td>impairment intraobserver reliability = 0.82; interobserver reliability = 0.79</td>
<td>partial correlation b/w total on this scale &amp; Quality of Life in Epilepsy −0.89 = −0.23 (p &lt; 0.0001) scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>International Classification of Functioning, Disability, and Health</td>
<td>impairment, activity, participation interrater reliability = 0.97 (47 National Health Service interviewers)</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>Karnofsky Performance Scale</td>
<td>activity test–retest reliability 0.80; Cronbach α = 0.85</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>Liverpool Seizure Severity Scale</td>
<td>impairment Cronbach α = 0.85 (0.003); &amp; mastery (0.003)</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>MMSE</td>
<td>impairment Cronbach α = 0.54–0.96; test–retest r = 0.80–0.95</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>National Hospital Seizure Severity Scale</td>
<td>impairment Cronbach α = 0.65 (removal of warnings before seizures &amp; automatisms increased α to 0.71)</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>NIH Stroke Scale</td>
<td>impairment interrater reliability (mean κ = 0.66–0.77), test–retest reliability (mean κ = 0.66–0.77)</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>Rankin Scale</td>
<td>activity interrater agreement K w = 0.75</td>
<td>scale-lesion size, r = 0.68; scale-outcome, r = 0.79</td>
</tr>
<tr>
<td>RNL Index</td>
<td>impairment coping Cronbach α &gt; 0.90 found when patients, significant others, or health professionals filled out form</td>
<td>mobility, ADL, instrumental activities, &amp; living arrangements highly associated w/ Rankin grades (Somers D range 0.60–0.74)</td>
</tr>
<tr>
<td>SF-36</td>
<td>activity Cronbach α = 0.77–0.94 in stroke patients</td>
<td>Pearson correlation = 0.67 on RNL daily functioning subscale compared w/ QOL index ADL item</td>
</tr>
<tr>
<td>SIP</td>
<td>participation test–retest reliability (r = 0.92) &amp; internal consistency high (r = 0.94)</td>
<td>physical function of SF-36 correlated strongly w/ FIM (r = 0.68) &amp; Expanded Disability Status Scale (r = 0.82)</td>
</tr>
<tr>
<td>Stroke Impact Scale</td>
<td>participation Cronbach α = 0.83–0.90</td>
<td>convergent validity: 0.60 correlation w/ Barthel Index &amp; 0.70 w/ global health score on the Rankin Scale</td>
</tr>
<tr>
<td>Stroke-Specific Quality of Life</td>
<td>participation in each of 12 domains Cronbach α was ≥ 0.73 test–retest reliability (r = 0.97)</td>
<td>Pearson correlation of 0.70 for SF-36 social function, 0.45 for SF-36 physical role overall score r² = 0.65 &amp; p &lt; 0.001 vs overall SF-36 score</td>
</tr>
<tr>
<td>Telephone Interview for Cognitive Status timed “Up-and-Go”</td>
<td>activity intrarater reliability (ICC) = 0.56 for cognitively impaired individuals</td>
<td>Pearson correlation 0.94 b/w Telephone Interview for Cognitive Status &amp; MMSE Spearman correlations (cognitively impaired) b/w timed “Up-and-Go” &amp; OARS ADL (–0.72), OARS instrumental ADL (–0.70)</td>
</tr>
</tbody>
</table>

(continued)
Wechsler Memory Scale is a widely used instrument for screening of memory function. 

**Scales for Measuring Mood Disorders.** Assessment of depression as a psychiatric impairment has been of interest in cerebrovascular outcomes studies because of the prevalence of detected mood disorders in stroke patients. Two popular scales are the Beck questionnaire, 22 which has been shown to be valid, reliable, and sensitive in physically healthy, psychiatrically ill patients. As noted by Wade, 185 the validity of this scale as a diagnostic measure of depression in physically disabled patients has not been established. The Zung Depression scale 280 was also developed to detect and measure depression in a population without stroke-related physical disabilities. The validity of these scales in the study of patients disabled by physical neurological impairment has not been established.

**Assessment of Activity-Level Outcomes for Cerebrovascular Disorders**

Activity-level outcomes are an important focus for patients, who are concerned about what they can do in terms of common tasks of both a physical and mental nature. Two scales of activity are frequently used as outcome measures: the GOS 64 and the Rankin Scale. 156 These two simple scales are widely used reporting measures that are easy to administer but are very crude, with low sensitivity for detecting between-group differences. The Rankin Scale, in particular, has been criticized for mixing impairment, activity, and participation aspects of recovery, but is most often categorized as a global measure of activity. 60,185 These measures are useful in large studies to permit rapid, easy-to-obtain, gross categorization of patients. When used alone, however, they have very little value in describing specific areas of outcome. In the case of newly developed therapies that are being tested, for example in a randomized trial, one may expect only an incremental benefit in outcome. Measures such as the GOS or Rankin Scale may not be able to detect a beneficial shift within categories. In the following sections we describe some available measures of both ADLs, referring to basic measures of self-care and physical activity, and the instrumental ADLs, which involve more complex cognitive and fine-motor tasks that allow a patient to live independently in the community.

**Activities of Daily Living.** The Barthel Index is the most widely used measure of ADLs. It is considered to have good reliability and validity but has been noted to be insensitive to smaller changes in functional status and to demonstrate ceiling effects 63 in patients with mild to moderate stroke as well as good-grade SAH. 99 This scale has been used to describe outcome after surgery for unruptured aneurysms and after surgery in older patients with ruptured lesions. 79

**Instrumental ADLs.** McDowell and Newell 230 have reviewed several instrumental ADL scales. A variety of these types of scales have been used in the stroke population, with less data about the use of these measures in SAH or ICH. Recently, the Functional Independence Measure has also been used as a global measure of activity in the poststroke population, 20,58,76,136,157,187 and it has also...
been used in the SAH population. The Functional Activities Questionnaire is a relatively recently designed scale that may have applicability in patients with dysfunction in cognitive domains after SAH. This is a 10-item instrument meant to be completed by a lay informant, spouse, or close friend, which focuses on cognitive functions required for independent living.145,146

Assessment of Participation and QOL-Level Outcomes in Cerebrovascular Disease

Measurements of participation and QOL are only now being developed in a stroke-specific fashion.62,191,192 “The Stroke Impact Scale” is a 64-item scale covering eight domains, including strength, hand function, ADLs, mobility, physical, memory/thinking, communication, emotion, and participation. This scale was developed in a population with mild-to-moderate stroke and has not been validated for use in severe cerebral infarction, SAH, or ICH. Discriminant validity was high for this instrument across four Rankin levels. A second instrument, the Stroke-Specific Quality of Life, is a 12-domain, 49-item scale.892 Domains were identified after detailed interviews with ischemic stroke survivors, and included mobility, energy, upper-extremity function, work/productivity, mood, self-care, social roles, family roles, vision, language, thinking, and personality. This scale is likewise untested in the population suffering severe stroke or in patients in whom an SAH or ICH has occurred.

In addition to disease-specific QOL instruments, more general instruments such as the SF-36 and SIP have been applied to stroke populations. The SF-36 has been used in populations of patients experiencing stroke7 as well as SAH.90 This scale enjoyed what has been described as a “meteoric rise to prominence” in the healthcare world due to a variety of factors; by the early 1990s more than 1 million forms were being administered per year.125 This standard of global health assessment has penetrated into a number of disease-specific domains treated by cerebrovascular surgeons. The SF-36 has 36 questions covering eight different domains of QOL, including physical functioning, role limitations due to physical health, bodily pain, social function, general mental health, role limitations due to emotional problems, vitality/energy or fatigue, and general health perception. The SF-36 has the advantage of being relatively quick to administer, using normative scoring with a mean of 50 and standard deviation of 10 for each of the two summary scales (physical health and mental health).

The SIP consists of 138 questions in 12 different areas of function. This examination may take as long as 1 hour to administer. This broad measure exemplifies the generic assessment approach and has been used in a wide variety of disease conditions and populations. It has been used to assess outcome after stroke,92 SAH,90 and perimesencephalic hemorrhage.92 The time needed for the examination has an impact on feasibility, yet the SIP is well established and is certainly an appropriate comprehensive measure.

The RNL Index has been used recently as a global outcome measure in stroke199 and SAH.99 The RNL Index is an 11-item QOL scale199 and has been validated for both in-person and telephone responses.109 The RNL can be completed in a matter of minutes. Reference values in elderly populations have also been obtained.174

Seizure Outcomes

Baker et al.18 have provided a comprehensive review of outcome assessment in epilepsy. They highlight the three scales that have been developed to quantify seizure severity: the Veterans Administration Seizure Frequency and Severity Rating Scale;47,150 the Liverpool Seizure Severity scale;21 and the National Hospital Seizure Severity Scale.199 Additional work has been performed to examine health-related QOL, with the most extensively tested instrument being the Epilepsy Surgery Index—55,183 This is a standardized 55-item, epilepsy surgery–specific, modular instrument composed of a 36-item generic core, the widely used SF-36,187 plus 19 additional items.

Disease-Specific Outcome Assessment

Extracranial Carotid Atherosclerotic Disease

Carotid atherosclerotic disease, by virtue of its prevalence and with established, successful interventions, has the highest number, greatest diversity, and best quality of outcomes analyses among the cerebrovascular disorders. The North American Symptomatic Carotid Endarterectomy Trial,14 the European Carotid Surgery Trial,15 and the Veterans Administration Cooperative Study showed that CEA improved outcomes in selected symptomatic patients with high-grade extracranial carotid artery stenosis for primary end points of death and stroke. Detailed systematic reviews of patient cofactors are now being presented163 that show the relative importance of factors that have an impact on postsurgical outcome, including transient monocular blindness versus cerebral transient ischemic attack, peripheral vascular disease, contralateral internal carotid artery occlusion, hypertension, ipsilateral external carotid artery stenosis, tandem stenoses, female sex, and older age. These cofactor analyses have permitted the extension of the original trial results to permit better preoperative risk stratification. Importantly, work has also been done to show that CEA improves the functional status of patients with symptomatic carotid stenosis better than medical treatment.96

Decision making regarding surgical treatment for asymptomatic carotid artery stenosis highlights the importance of cofactor analysis. Although an absolute reduction is seen in overall stroke rates for patients with greater than 60% stenosis, recommendations for surgical intervention presuppose a low perioperative stroke rate (< 3%) and a reasonable life expectancy (≥ 5 years). Given the more marginal benefit of CEA in asymptomatic disease, a cofactor assessment that delineates a higher surgical risk may serve to identify patients for whom the risk/benefit ratio for surgical intervention is unfavorable.

Additional advanced outcomes studies in carotid occlusive disease after CEA treatment are now focusing on patient satisfaction by comparing regional with general anesthetic,174 type of regional anesthesia block used,175 length of hospital stay issues,125,147 and nonneurological outcomes.143 Some efforts have also been made to examine QOL after CEA.80,184 and the economics of screening for asymptomatic116 and symptomatic stenosis.111 The neurosurgical community has advanced proposed online initiatives to monitor and track outcomes as well.92

C. E. Sanchez, C. S. Ogilvy, and B. S. Carter
A new generation of studies has been undertaken to assess the relative utility of carotid angioplasty and stent placement compared with CEA in terms of mortality rates and stroke morbidity outcomes (see review by Hanel et al.)]. Studies have been performed that supported either carotid angioplasty and stent placement or CEA or showed relative equivalency in terms of mortality rate and major morbidity risk. Several large studies are currently enrolling patients to further assess this question. After initial comparisons of stroke-related death and morbidity, it is anticipated that future studies will delve into more subtle secondary measures of outcome, including cognition, minor morbidity, and functional status. As the initial generation of stroke-related morbidity and mortality analyses comparing CEA and carotid angioplasty and stent placement are completed, one anticipates that there will be ongoing, detailed co-factor analysis that will be important for defining subgroups of patients in whom one or the other treatment is most applicable.

Intracranial Atherosclerosis and Carotid Occlusion

Most natural history and intervention outcomes studies for intracranial atherosclerotic disease and carotid occlusion have focused on stroke-related deaths and major morbidity. Most attention has been given to symptomatic rather than asymptomatic disease. Atherosclerotic narrowing of the intracranial vasculature increases in prevalence with patient age, and lesions demonstrate both progression and regression over time. In a recent review of retrospective studies examining deaths from intracranial atherosclerosis, the mortality rate was estimated to be in the 5 to 17% per annum range. The large prospective randomized Warfarin–Aspirin Symptomatic Intracranial Disease trial comparing aspirin and warfarin therapy showed a 19.7% stroke rate at 2 years for patients treated with aspirin. Patients with greater than 70% stenosis were at highest risk; they had a 2-year stroke rate of approximately 20%, compared with approximately 10% for those with less than 70% stenosis. In patients presenting with transient ischemic attack or stroke and concomitant intracranial internal carotid artery or middle cerebral artery stenosis, there are gains in functional status in the 1st year of follow-up care, but these gains are lost by 24 months of follow up and continue to decline thereafter. Prospectively determined stroke rates for patients with basilar, vertebral, or posterior inferior cerebellar artery stenosis have been reported, although the functional implications of these strokes have not. Outcomes studies have made an impact on therapeutic choices; the publication of the external carotid–internal carotid bypass trial effectively reduced the routine application of surgical bypass for the treatment of intracranial carotid stenoses. This was because of the failure to demonstrate reduction in stroke rates and later the demonstration of a worsened functional status in the patients treated with surgery. Numerous criticisms of this study have been presented, primarily centered on the question of patient selection and exclusions.

Carotid occlusion also confers an increased stroke risk, with patients who experience abnormal hemodynamics after occlusion demonstrating a stroke risk as high as 27% at 2 years, and 50% of patients demonstrating early symptomatology after occlusion. Renewed interest has led to the initiation of a new external carotid–internal carotid bypass trial (the Carotid Occlusion Surgery Study) to determine whether restoration of normal cerebral hemodynamics ipsilateral to carotid occlusion can reduce this stroke rate and improve stroke-specific QOL. As with prior analyses of surgical intervention, it will be important to analyze not only surrogates of outcome (for example, flow restoration and radiographically confirmed stroke rates) but also functional studies in which instruments that measure QOL are used.

Intracerebral Hemorrhage

Most natural history and intervention outcomes studies for ICH have focused on death and major stroke-related morbidity. A number of population-based studies have reported mortality estimates following ICH. A variety of factors have been analyzed in relation to mortality rates, including age, volume of hemorrhage, location of hemorrhage, presence of hydrocephalus, and increased systolic blood pressure at presentation. The level of consciousness at presentation is associated with the 30-day mortality rate in both medial (thalamic and caudate) and lateral (putaminal) hemorrhages. Multivariate analyses that include both hydrocephalus and volume of hemorrhage tend to include either variable in the final model; this is probably related to a high degree of correlation between these two variables.

Morbidity assessment for both natural history and intervention studies for ICH has generally focused on gross measures of activity such as the GOS. reported on Barthel Index Scores and NIH Stroke Scale scores in a randomized study of ultraearly surgical intervention after ICH. The International Surgical Trial in Intracerebral Hemorrhage reported by Mendelow and colleagues showed that surgical intervention within 24 hours of randomization offered no benefit in survival or prognosis-based indices for patients with lobar, basal ganglia, or thalamic ICH measuring more than 2 cm in diameter and a GCS score of 5 or more. In contrast, there was improvement in both survival and functional assessment according to modified Rankin Scale scores in patients receiving recombinant factor VII therapy after ICH.

Most ICH and natural history studies published to date have focused on death and major morbidity as the primary measures of outcome. In part, this relates to the high mortality rate associated with the disease and less focus on subtle cognitive improvements. As advances are made, particularly with medical management, increasing focus on documenting improvement in secondary measures such as cognitive performance and QOL outcomes will probably become more frequently reported.

Cavernous Malformations

Outcomes studies in cavernous malformations may be expected to require a multidimensional approach to delineate outcomes fully, including analysis of neurological impairment and seizure control as well as surrogate measures such as rehemorrhage rates. Death occurs relatively less frequently in patients with cavernous malformations compared with those who suffer ICH or SAH, although the rarity of this lesion and the infrequency of death make...
accurate estimation of a true case-fatality rate difficult. None of 32 patients younger than 40 years old who were admitted to a tertiary medical center died at presentation. No deaths occurred in two reported series of 110 patients. In a series of 100 brainstem cavernous malformations, three deaths were noted. Population-based case-fatality rates have not been reported.

Because death is less of an issue with cavernous malformations, investigators have reported on neurological outcomes and surrogate measures such as hemorrhage rates based on magnetic resonance imaging findings. Increasingly, neurological outcomes are the focus, as opposed to an earlier tendency to report the surrogate measure of hemorrhage. Porter et al. discussed the inherent difficulties of using hemorrhage rates as a surrogate for patient-based outcomes; in their study they noted a higher rate of cumulative neurological morbidity in patients with diencephalic or brainstem lesions measured with an author-defined scoring system for neurological outcome. Kupersmith et al. reported using the Rankin Scale outcomes in patients with brainstem cavernomas. Other natural history, surgical, or radiosurgical series have mostly used author-defined scales of neurological impairment or the GOS scale as a patient-based measure after treatment. Chang et al. reported the Karnofsky performance status of patients treated with radiosurgery. Cohen et al. used author-defined measures of seizure outcome to evaluate the role of lesionectomy for cavernous malformation treatment. Although there has been considerable use of both severity rating scales and QOL assessments in epilepsy surgery, a survey of recent articles in which either natural history or seizure outcomes were reported for medical, radiosurgical, or surgical interventions for cavernous malformations does not reveal the use of these instruments.

Arteriovenous Malformations

As for cavernous malformations, outcomes studies for AVMs require a diversity of outcome instruments. Neurological morbidity has been noted to be due to hemorrhage-related deficit, nonhemorrhage-related progressive deficit, seizure activity, and headaches. In two retrospective reviews the mortality rate was estimated to be approximately 1% per year after the diagnosis of a symptomatic AVM. Hemorrhage has been used as a surrogate measure for directly ascertained patient-based outcome measures. Hartmann et al. provided one of the more detailed analyses of the actual neurological impairment that can be expected with both an initial hemorrhage and repeated hemorrhage in patients referred to a tertiary medical center. Using the Barthel Index to assess neurological impairment, they examined patients at a mean follow-up time of 16.2 months after an initial hemorrhagic event. In their study, 93% of patients had a Barthel Index score of 100 after the initial hemorrhage and the same percentage achieved this score even after a subsequent hemorrhage. No patient died during the observation period. The authors also examined 59 cases of ICH from causes other than AVMs. In this latter group, the 1-year mortality rate was 27% and the mean Barthel Index score was 78.8, compared with a mean score of 98.3 in the AVM-related hemorrhages. Although Hartmann et al. suggested that AVM-related hemorrhages may yield less morbidity than previously thought, they did not assess other impairment-level outcomes such as cognition, or third-tier QOL outcomes after hemorrhage. Hartmann et al. were criticized for the relatively short follow-up period in their study and the potential referral bias to a tertiary care center, which could lead to deaths and devastating hemorrhages being missed. The complexity of the referral bias problem is highlighted by Hofmeister et al., who examined 1266 patients from three large AVM databases in Europe, Canada, and the US and found significant differences in these populations in terms of patient age and whether a hemorrhage as well as persistent and reversible neurological deficit had occurred.

Seizure outcomes after AVM surgery, radiosurgery, or embolization have been reported typically based on author-defined seizure outcomes. These reports have not used more standardized measures of QOL in this patient population.

Although widely used as a therapeutic modality, surgery has not been studied in direct comparison to natural history data in trial form to assess overall patient-based outcomes. In a number of surgical series, GOS scores or Rankin disability scores after surgical treatment of AVMs have been reported, and systems of cofactor analysis have been used to predict outcomes for surgical treatment of these lesions. Hartmann has analyzed cofactors and their relationship to disabling or nondisabling deficit as assessed using the Rankin scale after AVM surgery.

Subarachnoid Hemorrhage

The relevant outcomes in SAH are diverse, ranging from death to neurological impairment to cognitive impairment and decreased QOL. Significant nonphysical impairment after SAH has been the subject of increasing levels of scrutiny in recent years. It has long been noted that even in the patients assigned to the best-grade categories there exists a reduced capacity for functional activity and return to work.

Case-fatality rates for SAH have been reported in large-scale, population-based epidemiological studies. These studies have revealed important data regarding the incidence of SAH and case-fatality rates in the overall population of cases as well as those treated in the hospital. Smaller series have reported the lower mortality rate associated with the nonaneurysmal perimesencephalic variant of SAH. Population-based studies and multicenter hospital-based series have confirmed the association between older age and higher case-fatality rates. The initial level of consciousness is also associated with death in SA. Early surgical intervention has been associated with lower mortality rates in population studies. In a large randomized study (the International Subarachnoid Aneurysm Trial) in which a modified Rankin Scale was used, the mortality rate and dependency were compared between surgical and endovascular approaches to therapy for aneurysms associated with SAH.

Depression after SAH has been reported using the Beck Depression Inventory and the Zung Self-Rating Depression scale. Carter et al. suggested that depressive symptoms were correlated with the RNL's QOL.
scores. Cognitive outcomes after SAH have been assessed using a variety of neuropsychological test batteries. There is a paucity of data regarding the relationship between this outcome and measures of QOL in SAH, although this has been studied in patients with ischemic stroke.\textsuperscript{97,101,103,114,116} Cognitive impairment is known to affect the likelihood of nursing home placement in a geriatric rehabilitation population,\textsuperscript{55} but the level of cognitive impairment at which QOL scores begin to diminish is not well delineated. In several studies, analyses of higher cognitive functions after SAH have been used.\textsuperscript{117,129} Studies have been completed using the Wechsler Adult Intelligence Scale,\textsuperscript{68,140} the Wechsler Memory Scale,\textsuperscript{68,140} the MMSE,\textsuperscript{125} and the Cambridge Neuropsychological Test Automated Battery.\textsuperscript{125}

Neuropsychological outcomes have been used to assess treatment group differences in SAH. Koivisto et al.\textsuperscript{105} randomized 109 patients with aneurysmal SAH to either surgical or endovascular treatment. The mortality rate was 17% according to assigned treatment in either group. At 1-year posthemorrhage, there were no significant differences between the two groups in terms of the three verbal subtests of the Wechsler Adult Intelligence Scale–Revised. Nonverbal intelligence as estimated using the picture completion and block design tests of the Wechsler Adult Intelligence Scale–Revised also showed no between-group differences.

Dombrovsky et al.\textsuperscript{57} tested 80 patients with SAH by using the Functional Independence measure on admission and at discharge from inpatient rehabilitation. Thirty-two of the 80 patients were tested again 2 years later and found to have achieved ongoing improvement in physical functioning on the Functional Independence Measure.\textsuperscript{37} However, 40% of these patients remained impaired when assessed with the Telephone Interview for Cognitive Status measure. Carter et al.\textsuperscript{39} determined Barthel Index scores in a group of patients with SAH who were conscious and conversant at admission (Hunt and Hess Grades I–III) and found a significant ceiling effect of the Barthel Index in this group of patients. Activity impairment as assessed using this instrument and reduction in return-to-work rates have been noted in two series of patients.\textsuperscript{39,106,137,138}

Several QOL instruments have been used to report outcomes after SAH, including the SF-36;\textsuperscript{53,77,90} the SIP\textsuperscript{99} and a derivation thereof, the Aachen Life Quality Inventory; and the RNL Index.\textsuperscript{39} In all three of the studies in which the SF-36 was used, the investigators reported very significant reductions in both the “Role Limitations due to Emotional Problems” and the “Role Limitations due to Physical Problems” dimensions of the SF-36. The two questions on the SF-36 that correspond to these areas inquire whether the patient has reduced the amount of time spent or accomplished less at work or other activities due to physical or emotional problems. Notably, the 10-item “Physical Functioning” subcomponent of the SF-36 was always less reduced than the “Role Limitations due to Physical Problems” score, suggesting that ADLs as assessed by the former component are relatively less impaired than those aspects of physical health that would allow a person to assume his or her former roles. In a multivariate analysis, Carter et al.\textsuperscript{39} showed that both physical limitation and depression were correlated with the RNL outcome measure, with 55% of the patients classified in Hunt and Hess Grades I through III reporting an optimal QOL. The International Subarachnoid Aneurysm Trial\textsuperscript{132} also reported on seizure outcomes, comparing surgical and endovascular therapies after SAH.\textsuperscript{132}

Unruptured Intracranial Aneurysms

Brennan and Schwartz\textsuperscript{31} have recently reviewed the literature on unruptured intracranial aneurysms. This disease category represents a special case for aneurysm-related outcomes analysis. Because of the excellent condition of the patients before treatment, outcomes analyses are likely to be most useful in focusing on the probable deficits in a minimally impaired patient population. Analysis of ADLs will demonstrate ceiling effects because most patients will do well; however, systems such as the Barthel Index will be useful in determining the percentage of patients who convert from an independent to dependent functional status after treatment. Because this treatment regimen may offer distinct cost/benefit alternatives (for example, reduction in deaths from aneurysm rupture at a cost of partial degradation of cognitive function), it will be useful for future studies to extend cofactor analyses to determine subtypes of patients who are at high risk for poor outcomes.

The International Unruptured Intracranial Aneurysm study\textsuperscript{39} paved the way for the first large-scale examination of the effects of treating unruptured intracranial aneurysms. The coupling of low hemorrhage rates with an unexpectedly high rate of negative treatment effects has raised questions about the utility of intervening in certain groups of patients with unruptured aneurysms. In the first report based on this analysis, the authors prospectively analyzed the 1-year outcome for 798 patients (Group 1) who underwent surgical (83%) or endovascular (17%) treatment for an unruptured aneurysm and for 197 patients (Group 2) with prior SAH who underwent craniotomy (93%) or endovascular (7%) treatment for a second, unruptured aneurysm. The total 1-year mortality rate in the Group 1 patients was 2.3%, with a Rankin Scale score of less than 2 in 6.6% of patients, and an MMSE score of less than 24 or a Telephone Interview for Cognitive Status score of less than 27 in 8.7% of patients, representing a severe cognitive deficit based on population norms.\textsuperscript{48} This initial report must necessarily be viewed as an important preliminary study that raises several sets of important questions. The next round of data that should be forthcoming regarding unruptured aneurysms will be subgroup-specific outcomes that involve earlier, cruder measures such as the GOS or the Rankin disability scale.

Raaymakers\textsuperscript{155} examined 18 patients with detailed preangiography, presurgery testing including the SIP, SF-36, Barthel Index, and Rankin Scale. All patients were examined again at 3 months and 1 year postoperatively. Before angiography and surgery, all patients had optimal Rankin Scale and Barthel Index scores. By 1 year postprocedure, all patients had an optimal Barthel Index, whereas a subset of patients had a decrement in the SIP.

Subsequent to this, we anticipate that detailed neuropsychological analyses and QOL assessments will be forthcoming for both treated and untreated patients with an unruptured intracranial aneurysm.\textsuperscript{132}

Decompressive Surgery for Major Stroke

The resurgence of interest in the use of decompressive...
procedures for major stroke offers several opportunities for outcomes analysis. It has been suggested that decompressive surgery for major supra- and infratentorial 2,189 and intracerebral hemorrhage reduces mortality rates when compared with natural history or historical controls. In two large case series in recent years 40,166 the Barthel Index has been used to characterize physical recovery after supratentorial hemicraniectomy. Interestingly, the two studies, one performed in Germany and one in the US, defined a very similar level of disability in survivors, with a mean Barthel Index of 68 in both reports. Carter et al. 40 also used the RNL Index to examine the QOL of survivors and found significant decrements in QOL as assessed by this measure. Kelly et al. 184 showed substantial improvement in the Functional Independence measure for patients undergoing rehabilitation for cerebellar stroke, including 15 patients treated with suboccipital craniectomy.

Spinal Vascular Malformations

Spinal dural arteriovenous fistulas have been the subject of a number of series examining outcome, 2,189 although in these series outcome has been simply reported as improved, stable, or worse. As discussed in Oder et al., Behrens and Thron 13 used the Barthel Index as well as their own pain and impairment indices to assess pre- and posttreatment disability in 21 patients undergoing surgical occlusion of a fistula. The mean posttreatment Barthel Index of 86 was significantly improved compared with the pretreatment score of 70. This study, in which patients acted as their own controls, strongly confirmed the role of early occlusion of the fistula in improving motor status and functional ability in these individuals.

Conclusions

The range of outcomes instruments available and the application of these instruments to problems in cerebrovascular surgery are increasing. Several recommendations can be proposed for investigators designing outcomes studies in cerebrovascular surgery.

First, for conditions or studies in which exploratory hypotheses or preliminary outcome assessments are being generated, strong consideration should be given to inclusion of a measure of activity as a primary outcome measure. Coarse measures such as the GOS or Rankin Scale can be supplemented with a more specific ADL measure such as the Barthel Index. In populations in which the Barthel Index would be expected to show a ceiling effect, such as the Barthel Index. In populations in which the Functional Activities Questionnaire or the Functional Independence Measure would also be appropriate. These latter measures will allow for more complex cognitive and fine-motor assessment. As secondary outcome measures, assessments at the level of impairment will allow for stratification of the outcomes analysis based on severity of deficit; this will aid in the interpretation of studies across different populations. In the case of cerebral ischemia, motor, mood, and cognitive impairment have complex interrelationships, and each should be considered as a potential outcome measure.

Second, retrospective studies with telephone-based fol-

low up should use measures that have been validated for telephone interviews, postal self-assessment, and/or proxy respondents. This type of validity has been shown for such measures as the Barthel Index, 109 the SF-36, 188 the RNL Index, 109 and the Telephone Interview for Cognitive Status. This offers the opportunity for measuring outcomes at several levels in a situation in which the interview has not been conducted face to face.

Third, randomized trials of specific interventions are most likely to yield evidence of effect when specific and sensitive outcome measures are used. Randomization implies that the outcome is less than certain. Coarser measures of outcome may obscure a within-category beneficial effect. The Rankin Scale and GOS may be too coarse to detect an intervention benefit.

Fourth, data collection should include baseline characteristics that are known to influence natural history and could confound outcomes interpretation.

Fifth, the development of specific impairment assessments for certain conditions may be appropriate in cases in which little validation or adaptation of existing instruments has occurred. Guidelines for creating a cerebrovascular health measure have been reviewed elsewhere. 60,128

References


C. E. Sanchez, C. S. Ogilvy, and B. S. Carter

Neurosurg. Focus / Volume 22 / March, 2007

Unauthenticated | Downloaded 04/19/22 01:26 PM UTC
Outcomes studies in cerebrovascular neurosurgery


C. E. Sanchez, C. S. Ogilvy, and B. S. Carter
Outcomes studies in cerebrovascular neurosurgery

120. Manning C: Carotid endarterectomy path cuts LOS below two days. Hosp Case Manag 5:199–202, 1997
132. Moore L, Lavoie A, Camden S, Le Sage N, Sampalis JS, Ber-


174. Steiner A, Raube K, Stuck AE, Aronov HU, Draper D, Rubenstein LZ, et al: Measuring psychosocial aspects of well-
Outcomes studies in cerebrovascular neurosurgery


__________________________


Address reprint requests to: Bob S. Carter, M.D., Ph.D., Yawkey 9026, Massachusetts General Hospital, 32 Fruit Street, Boston, Massachusetts 02114. email: bcarter@partners.org.