Infarct Volume as a Surrogate or Auxiliary Outcome Measure in Ischemic Stroke Clinical Trials

Jeffrey L. Saver, MD; Karen C. Johnston, MD; Daniel Homer, MD; Robert Wityk, MD; Walter Koroshetz, MD; Laura L. Truskowski, PhD; E. Clarke Haley, MD; for the RANTTAS Investigators

Background and Purpose—Reduction in infarct volume is the standard measure of therapeutic success in animal stroke models. Reduction in infarct volume has been advocated as a biological surrogate or auxiliary outcome measure for human stroke clinical trials to replace or supplement deficit, disability, and global clinical scales. However, few studies have investigated correlations between infarct volume and clinical end points in acute ischemic stroke patients.

Methods—CT scans at days 6 to 11 were acquired prospectively in 191 fully eligible patients enrolled in the Randomized Trial of Tirilazad Mesylate in Patients With Acute Stroke (RANTTAS). Patients were enrolled within 6 hours of onset of stroke in any vessel distribution. Infarct volume was measured by operator-assisted computerized planimetry.

Results—One hundred thirty-two patients had visible new supratentorial infarcts, with median infarct volume of 28.0 cm$^3$ (interquartile range, 9.0 to 93.0 cm$^3$). Fifty-nine patients had no visible new infarct. Correlations with standard 3-month outcome scales and mortality were as follows: Barthel Index, \( r=0.43 \); Glasgow Outcome Scale, \( r=0.53 \); National Institutes of Health Stroke Scale, \( r=0.54 \); mortality, \( r=0.31 \). For visible infarcts alone, correlations were as follows: BI, \( r=0.46 \); GOS, \( r=0.59 \); NIHSS, \( r=0.56 \); mortality, \( r=0.32 \).

Conclusions—Subacute CT infarct volume correlates moderately with 3-month clinical outcome as assessed by widely used neurological and functional assessment scales. The modesty of this linkage constrains the use of infarct volume as a surrogate end point in ischemic stroke clinical trials. (Stroke. 1999;30:293-298.)

Key Words: cerebral infarction ■ clinical trials ■ stroke assessment ■ stroke outcome ■ tomography, x-ray computed

Randomized clinical trials employ a variety of end points to determine the effectiveness of a novel agent. For definitive phase 3 trials, the primary end point should be a clinical outcome of direct relevance to the patient. However, important clinical outcomes are often infrequent and delayed, and phase 3 trials consequently may be time-consuming and expensive. In phase 2 trials, the goal is rapid, inexpensive screening of new agents and dosages for biological activity and safety. For these purposes, using clinical outcome measures alone may mandate impractical sample sizes. Clinical trialists and statisticians therefore have sought intermediate, surrogate end points.

A surrogate end point, to quote Temple’s definition from among many available, “is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.” The concept of a surrogate end point has been precisely made operational with the use of statistical formulas. Ideally, surrogate end points should be simple to ascertain, easily quantifiable, inexpensive, and tightly correlated with targeted clinical outcomes. A related concept is that of an auxiliary end point. An auxiliary end point is an intermediate biomarker for a true clinical end point that, by demonstrating a similar response to an intervention, supplements or strengthens standard analysis of true clinical end point data.

Phase 3 clinical trials of therapies for acute ischemic stroke currently use as primary end points scores at 3 to 6 months on rated measures of activities of daily living, disability, neurological deficit, and/or global outcome. Drawbacks of these instruments include acceptable but not excellent interrater reliability, insensitivity to fine variations in treatment effectiveness, the costs of maintaining patient follow-up over a 3- to 6-month period, and extreme variability. In contrast,
measurement of infarct volume is the standard gauge of therapeutic efficacy in animal stroke models. Decrease in infarct volume has been suggested as an appropriate surrogate or auxiliary outcome measure for human stroke trials.\textsuperscript{8,9} However, few reports have analyzed correlations between infarct volume and clinical end points in stroke trial populations.

We studied correlations between infarct volume and location and clinical outcomes in the Randomized Trial of Tirilazad Mesylate in Patients With Acute Stroke (RANTTAS).

**Subjects and Methods**

RANTTAS was a multicenter, randomized, double-blind, vehicle-controlled trial to evaluate the safety and efficacy of intravenous tirilazad mesylate in patients with acute ischemic stroke.\textsuperscript{10} Tirilazad is a nonglucocorticoid 21-aminosteroid that inhibits lipid peroxidation in vitro and reduces infarct volume in animal models of focal ischemia. Patients were enrolled within 6 hours of ischemic stroke in any vessel distribution.

Cranial CT scans were performed on days 6 to 11 in a prospective random sample of 50\% of fully eligible subjects. Noncontrast scans were obtained according to a standardized protocol that specified 5-mm slice intervals. Volume of cerebral infarction was measured centrally by an investigator blinded to treatment group using computerized planimetric techniques.\textsuperscript{8} Localization analysis was performed centrally by investigators blinded to treatment group, using a checklist of anatomic structures encompassed by the lesion. Involvement or sparing was noted for the following 9 regions: frontal, parietal, temporal, occipital, corona radiata/internal capsule, basal ganglia, thalamus, cerebellum, and brain stem.

Patients were included in this study if they fulfilled the following criteria: technically adequate CT scans were performed in the day 6 to 11 window, with the films received at the coordinating center; infarct location was determined to be hemispheric; no pre-stroke disability was present (determined by history obtained at baseline entry into the trial); and 3-month outcome evaluations were obtained between days 76 and 106 after entry, or patients died before that window. Patients with brain stem/cerebellar strokes were excluded from this analysis because infarcts in this location are not reliably detected by CT due to bone artifacts. Patients with pre-stroke disability, from either prior stroke or other causes, were excluded because of the potential confounding effects on the outcome disability measures. Patients who had their outcome evaluations performed outside the day 76 to 106 window were therefore excluded from this study because of the potential confounding effects of changing neurologic and functional status in patients recovering from stroke.

Testing for a treatment effect with the use of the Kruskal-Wallis test demonstrated no significant effect of treatment assignment on baseline CT. The median NIHSS score on entry was 10 (IQ range, 5 to 16).

**Results**

One hundred ninety-one patients met the inclusion criteria for this study. Of 256 CT studies obtained, 3 were technically inadequate, 17 had clinical brain stem/cerebellar infarcts, 25 were performed outside the day 6 to 11 window, and 20 had prestroke disability. Of the 191, 20 patients died during follow-up and were therefore excluded from the BI and NIHSS analyses. Twenty-one patients had their 3-month follow-up scores performed outside the day 76 to 106 window and were therefore excluded from the BI, GOS, and NIHSS analyses. An additional 17 patients did not have the NIHSS performed at follow-up, and 2 more had 1 or more missing subscores at follow-up. The total number of subjects for the mortality analyses was therefore 191; for the GOS, n = 170; for the BI, n = 150; and for the NIHSS, n = 131.

Among the 191 study patients, 57\% were male, 43\% were female, and the mean\(\pm SD\) age was 69\(\pm\)12 years. The RANTTAS qualifying stroke was the first-ever stroke in 64\%, while 15\% had prior clinical strokes, 16\% had prior silent strokes (evident on baseline CT only), and prior history information was unavailable in 1\%. Among those with a clinical history of prior stroke, 61\% had an old stroke visible on baseline CT. The median NIHSS score on entry was 10 (IQ range, 5 to 16).

On the day 6 to 11 CT, 132 patients had visible new infarcts and 59 had no visible infarcts. The overall median infarct volume was 10.5 cm\(^3\) (IQ range, 0 to 65.0 cm\(^3\)). For the 132 visible infarcts, the median infarct volume was 28.0 cm\(^3\) (IQ range, 9.0 to 93.0 cm\(^3\)). The correlations between infarct volume and 3-month outcome are shown in Table 1. Correlation coefficients ranged from 0.31 for mortality to 0.54 for the NIHSS. Limiting the analysis to just the 132 patients with new visible infarcts did not substantially change the results. Imputing worst outcome scores on the BI and the NIHSS for patients who died before 3-month follow-up also did not significantly alter correlation values (data not shown). Figures 1 to 4 graphically illustrate the relationships between infarct volume and clinical outcome.
Analyses of infarct lateralization and pure subcortical, pure cortical, and mixed subcortical-cortical subgroups failed to yield significant improvements in the correlations (data not shown). Analyses of infarct localization by neuroanatomic region were not sufficiently powered to detect major localization effects (1 data table filed with the National Auxiliary Publications Service).

Table 2 shows the odds ratios and 95% CIs for a dichotomized favorable outcome associated with a unit increase in infarct volume. None of the CIs include 1.0. Table 3 shows the median infarct volumes for favorable and unfavorable outcomes among the different outcome measures. Patients with favorable outcomes had median infarct volumes <10 cm³, while patients with unfavorable outcomes (except for the NIHSS, which scored anything other than complete or nearly complete recovery on the neurological examination as unfavorable) had median infarct volumes >100 cm³.

Discussion

In this sample from a large, multicenter clinical trial, infarct volume correlated moderately with standard clinical measures of stroke outcome. Correlation coefficients were in the 0.43 to 0.54 range for infarct volume alone versus 3 common clinical outcome measures after stroke: the BI, NIHSS, and GOS.

Comparable prior studies are sparse and contradictory. Several small series demonstrated no significant statistical correlation between CT lesion volume and clinical measures (BI, NIHSS, Rankin Disability Scale). Other, generally larger, series did show correlations between subacute CT lesion size and clinical scales (NIHSS, Rankin Disability Scale, Oxford Disability Scale, aphasia severity scale). Compared with these earlier investigations, the RANITAS data set is unique in size, geographic diversity, and actual participation in an acute stroke clinical trial. Our findings consequently are more likely to approximate the true relationship between subacute CT infarct volumes and 3-month clinical measures, especially in clinical trial settings.

Are correlation coefficients in the 0.31 to 0.54 range adequate to endorse CT infarct volume as a potential surrogate end point for clinical measures of outcome? This linkage
clearly falls short of an ideal biomarker and is insufficient to allow CT infarct volume to serve as a replacement end point for clinical markers in pivotal phase 3 efficacy trials, for which false-positive and false-negative rates must be low, usually 2.5% to 10%. However, precedent suggests that this degree of linkage may be sufficient to permit adopting CT infarct size as an auxiliary measure of outcome in assessing phase 3 trials. The most influential auxiliary outcome measure in neurological pharmaceutical trials in the 1990s has been MRI-measured plaque volume in multiple sclerosis. However, the correlation between MRI plaque volume and the expanded Kurtzke Disability Scale is only 0.23 to 0.33, substantially less than the correlations we observed between CT infarct volume and stroke disability and activity of daily living scales.

An important factor potentially attenuating the relation between infarct size and clinical outcome measures is lesion location. The site of injury is a consequential additional variable determining the extent of functional deficits from stroke. The effect of location reflects human brain organization into regionally distributed, large-scale neural networks subserving specialized functions, including motor control and language, with disparate impact on clinical outcome. In the RANTTAS data set, limited cell size in individual neuroanatomic regions limited exploration of and adjustment for the effects of lesion localization.

Several additional factors besides lesion location likely weaken correlations between infarct volume and clinical outcome. Confounding variables include the following: (1) patient age; (2) patient sex; (3) baseline cerebral atrophy; (4) prior symptomatic and silent infarcts; (5) prior nonvascular cerebral insults; (6) “fogging” effect on 1-week CT; (7) variable development of medical complications of stroke, including pneumonia, deep venous thrombosis, and pulmonary emboli; (8) interindividual variations in neuroplasticity and recovery capacity; (9) variable exposure to pharmacological agents that secondarily enhance or inhibit neuroplasticity and recovery; and (10) degree of family and community social supports.

Figure 3. Box and whiskers plot of Infarct volume vs GOS. GR indicates good recovery; MD, moderate disability; and SD, severe disability. The horizontal line indicates the median value, the box delineates the middle 50% of values (the IQ range), the whiskers demarcate values within 1.5 IQ ranges, and the small box with plus symbol indicates observations >3 IQ ranges below the first quartile or above the third quartile.

Figure 4. Box and whiskers plot of infarct volume vs mortality. Marking conventions are as detailed for Figure 3.
In our population, day 6 to 11 CT failed to visualize an infarction in 33% of patients. This detection failure rate is consistent with those as high as 23% to 59% observed in prior studies.7-26 In part, this reflects intrinsic limitations in the CT technique, including limitations on spatial resolution, fogging effect of luxury perfusion obscuring hypodensity, and skull x-ray beam hardening.30 An additional factor in our population is the effect of hyperacute patient selection. A proportion of patients enrolled within the first 6 hours of ischemic symptom onset may show early subsequent improvement and clinically exhibit no or minimal residual deficit at follow-up. These patients are likely to have smaller volume infarcts than patients with significant deficits at 24 hours. Correlation coefficient values between infarct volumes and clinical outcomes were not substantially different when analysis was confined only to those who had a visible infarct, compared with all eligible patients. Thus, including the cases with no visible infarct did not significantly alter our correlation results. However, the lack of visible CT lesions in a substantial minority of patients, a “floor effect,” is an important constraint on the use of CT measures of lesion volume as outcome markers.

Two other potential limitations of our study may be noted. We obtained CT scans over a relatively broad day 6 to 11 window. Mass effect from cytotoxic edema may be variably resolving during this period and may be expected to be more substantial in earlier than later scans, a possible confounding effect. To address this issue, we performed separate correlation analyses for early CT window (days 6 to 8) and late CT window (days 9 to 11) patients (data not shown). There was mild improvement in the ability of CT to predict mortality, but there were no major differences for BI, GOS, and NIHSS correlations. Second, approximately 10% to 15% of patients receiving CT scans had day 76 to 106 data missing, their final trial visit having occurred before or after this window. To analyze whether including these patients would alter our findings, we performed a last observation carried forward analysis for the BI, GOS, and NIHSS (data not shown). Correlation coefficients with infarct volume did not materially change. Thus, neither the use of the day 6 to 11 CT window nor the day 76 to 106 clinical outcome window substantially lessened correlations between CT infarct volume and clinical outcome.

The modesty of the correlation between CT infarct volume and standard measures of clinical outcome suggests that CT volume could be considered a somewhat useful but imperfect auxiliary outcome marker for ischemic stroke studies. New imaging techniques may provide better early biological markers of treatment effectiveness. For example, T2-weighted MR measures have greater spatial resolution and sensitivity to ischemic injury than CT attenuation. Few series have compared conventional MR measures of infarct volume and clinical end points, but the largest such study supports the hypothesis that MR measures may more closely correlate with functional outcome. Among 50 patients, MR scans performed 2.6 to 44.8 weeks after onset were compared with NIHSS scores obtained the same day and BI scores obtained 12 weeks after onset. T2-weighted infarct volume measures correlated somewhat more closely with BI and NIHSS scores than did CT infarct volumes in the RANTTAS data set.31 Even more promising as candidate auxiliary outcome measures are novel MRI diffusion and perfusion techniques that allow early patient-specific volumetric assessment of the region at ischemic risk and late determination of the volumes of infarcted and salvaged tissues.32,33 These methods, however, will require rigorous evaluation before their acceptance as adequate auxiliary or surrogate clinical trial outcome measures.

It is important to note that imaging markers, no matter how accurate, can never fully replace clinical assessments of long-term outcome. Dissociated effects of treatments on lesion volume and clinical outcome can occur through a variety of mechanisms. For example, an effective acute neuroprotective agent that had severe but delayed neurotoxic effects would yield favorable early imaging outcomes but unfavorable long-term clinical outcomes. Conversely, a neuroplasticity-enhancing agent that did not alter initial injury but improved functional recovery in uninjured regions would yield unfavorable early imaging results but favorable long-term clinical outcomes. Accordingly, clinical outcome scales should continue to serve as the primary end points in definitive, phase 3 randomized clinical trials of novel therapies for acute ischemic stroke.

**TABLE 2. Influence of Infarct Volume on Favorable BI, GOS, and NIHSS Outcomes and on Mortality**

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Odds Ratio (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>1.019 (1.010–1.028)</td>
</tr>
<tr>
<td>GOS</td>
<td>1.026 (1.016–1.035)</td>
</tr>
<tr>
<td>NIHSS</td>
<td>1.050 (1.028–1.073)</td>
</tr>
<tr>
<td>Mortality</td>
<td>0.988 (0.982–0.993)</td>
</tr>
</tbody>
</table>

*A odds ratios are for each 1 mm³ decrease in infarct volume and are adjusted for age, sex, prior stroke, and old infarct on CT scan.

**TABLE 3. Median Infarct Volumes in Patients With Favorable and Unfavorable Outcomes**

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Response</th>
<th>Infarct Volume, cm³</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>Favorable</td>
<td>3.5 (0–25.7)</td>
</tr>
<tr>
<td></td>
<td>Alive with</td>
<td>100.2 (25.1–156.9)</td>
</tr>
<tr>
<td></td>
<td>unfavorable</td>
<td></td>
</tr>
<tr>
<td>GOS</td>
<td>Favorable</td>
<td>3.2 (0–21.7)</td>
</tr>
<tr>
<td></td>
<td>Unfavorable</td>
<td>108.0 (25.1–181.4)</td>
</tr>
<tr>
<td>NIHSS</td>
<td>Favorable</td>
<td>0.7 (0–67.6)</td>
</tr>
<tr>
<td></td>
<td>Alive with</td>
<td>46.4 (1.9–108.7)</td>
</tr>
<tr>
<td></td>
<td>unfavorable</td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>Surviving</td>
<td>7.4 (0–49.2)</td>
</tr>
<tr>
<td></td>
<td>Dead</td>
<td>123.9 (19.4–213.4)</td>
</tr>
</tbody>
</table>

*Values for infarct volume are median (IQ range).
Infarct Volume as a Stroke Trial Surrogate Outcome Measure

NY 11552-2664. One explores the influence of involvement of 9 neuroanatomic regions on clinical outcome; the other lists sample sizes required for clinical trials using different clinical and CT end points, based on measure variability observed in the RANTTAS trial.

References
Infarct Volume as a Surrogate or Auxiliary Outcome Measure in Ischemic Stroke Clinical Trials
Jeffrey L. Saver, Karen C. Johnston, Daniel Homer, Robert Wityk, Walter Koroshetz, Laura L. Truskowski and E. Clarke Haley
for the RANTTAS Investigators

Stroke. 1999;30:293-298
doi: 10.1161/01.STR.30.2.293

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/30/2/293

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/