the intra-arterial group, compared with one of the six animals in the intravenous group, and the infarction rate was identical with the infarction rates in the two control groups. It would have been useful for Russell et al to quantify ischemic injury and correlate these changes with the recanalization data. The importance of this point, as it concerns outcome, is stressed by the results for those animals that received a rapid intravenous (square-wave) infusion with high-dose t-PA (10 mg/kg); patency rates by Doppler ultrasound were 100% at 2 and 18 hours, but pathological evaluation demonstrated infarction in 3% and intracerebral hemorrhage in 25% of these animals.

We quantified the volume of ischemic injury using triphenyltetrazolium chloride and found the percent of whole brain damage after either intravenous (3.4±2.6%) or intra-arterial (4.6±4.1%) infusion to be very small, and significantly smaller (P<.0001) than in the control (20.1±4.6%) animals. The concerns expressed by Clark et al concerning the possibility of type II error and the importance of dosing requirements in determining the efficacy of thrombolytic agents were thoroughly addressed in the "Discussion" section of our article. When we compared the infarct size in the intra-arterial and intravenous treatment groups, the probability that the difference between them was due to chance was very high (P=.79). This probability, combined with the similarity of the variances for the data in the two treatment groups (and the control group), reduces the risk of a type II error. In the same regard, it should be noted that Russell et al use the results from two small treatment groups (n=6 for each group) to argue that intra-arterial therapy is significantly (P<.05) more efficacious than intravenous therapy.

We feel that the difference in outcome criteria makes any direct comparison between these two studies most difficult. In addition, the differences in the type and age of thrombus used for embolization—we used 24-hour-old arterial thrombus and Russell et al used venous thrombus aged for 2 hours—add to the difficulty of comparing results.

Finally, we would agree with Clark et al that new techniques for the supraselective delivery of thrombolytic therapy may have advantages over previously available arterial or intravenous administration protocols for the lysis of clot. However, as pointed out in the editorial comment that followed our article, the major drawback of these techniques is the necessary delay imposed by the technical aspects of performing the angiogram and supraselective catheterization. As Dr Clarke Haley commented, "[i]t is possible that such delays might reduce both the benefit and the safety of the procedure to such an extent as to outweigh its advantages." The lack of generalized availability of these specialized techniques might further negatively influence or delay treatment.

Again, we would like to thank the writers for their comments. We would certainly agree with them that it is important to exercise caution when extrapolating the results from any animal study to a clinical situation. Nevertheless, the rationale for initiating thrombolytic therapy as soon as possible after the onset of stroke is strong, and intravenous infusions ensure the most rapid initiation and widest availability of treatment. Clinical trials testing this hypothesis are already in progress.

Joseph M. Zabramski, MD
Division of Neurological Surgery
Barrow Neurological Institute
Phoenix, Ariz

Mark G. Hamilton, MDCM, FRCS(C)
Division of Neurosurgery
Alberta Children's Hospital
Calgary, Canada

References


Sample Size Calculations for Clinical Trials Using the NIH Stroke Scale

To the Editor:

The National Institutes of Health Stroke Scale (NIHSS) is a validated and reliable measure of stroke-related neurological impairments that is being used with increasing frequency as the primary measure in therapeutic trials. The recent report by Wityk et al6 provides invaluable longitudinal NIHSS data for stroke patients similar to those who are generally included in interventional randomized control studies (eg, patients with carotid distribution ischemic stroke, first evaluated within hours of the onset of symptoms, and with an initial deficit of at least 4 NIHSS points). Final follow-up evaluations were performed by Wityk et al at a mean of 44 days after stroke (at a time when the most rapid period of spontaneous recovery has been completed). These data permit calculation of the sample sizes for various effect sizes and serve as an important reference for the interpretation and planning of studies using the NIHSS.

Assuming that data are normally distributed, the sample size needed for a given difference in mean scores between groups to be significant is

\[
\frac{Z\alpha + Z\beta}{2} \sqrt{\frac{2\sigma^2}{a-b}}
\]

where \(\alpha\) = probability of making a type I error, \(\beta\) = probability of making a type II error, \(\sigma\) = standard deviation, and \(\delta\) = difference in means between control and experimental groups.4

The standard error of the mean at the time of final follow-up for all 50 patients in the study was 1.3 NIHSS points.5 That for the 27 patients with middle cerebral artery distribution (MCA) embolic stroke was 2.1 points and that for the 14 patients with lacunar stroke was 1.1 points. The standard deviations were 9.2, 10.9, and 4.1 points, respectively. Empirically, a 4-point difference in the NIHSS score has been considered clinically significant.3,4 Setting \(a = .05\) (two-sided; \(Z_{1.642} = 1.96\)) and power at 80% (\(\beta = .2\); \(Z_{0.84} = 1.645\)), for a 4-point difference between treatment and control groups at the time of follow-up to be statistically significant, 71 patients would be required in each treatment arm if patients with both MCA and lacunar strokes were studied and 83 patients would be required in each treatment arm if only patients with presumed MCA strokes were included. This assumes no mortality and complete data on all patients. Because the initial severity of the deficit is correlated with final outcome, there could be no significant baseline differences between the groups. Fewer patients would be required in each treatment arm if larger differences in NIHSS scores are considered clinically significant.

A given NIHSS score can describe very different functional impairments (the scale is not weighted and measures a variety of impairments). Therefore, analysis with parametric statistics is likely not appropriate. A more suitable method of analyzing the data is to dichotomize the outcomes and compare the proportions of patients in treatment and control groups who achieve a given degree of improvement. Wityk et al6 found that 19 of 37 patients (51%; 95% confidence interval, 35%–65%) with an initial NIHSS score of at least 4 points had a major improvement at the time of follow-up (assuming that a 4-point change was indicative of a major neurological improvement). For an effect size of 10% to be detected (ie, to increase the percentage of patients achieving a major neurological improvement from 50% to 55%) with \(a = .05\)
TABLE 1. Sample Sizes for All Patients With Initial NIHSS Score ≥4 (α=.05)

<table>
<thead>
<tr>
<th>Response in Treatment Group</th>
<th>n (Per Group)</th>
<th>Power 80%</th>
<th>n (Per Group)</th>
<th>Power 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% “major improvement” in control group</td>
<td>1605</td>
<td>2135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65%</td>
<td>182</td>
<td>239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35% “major improvement” in control group</td>
<td>1511</td>
<td>2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45%</td>
<td>395</td>
<td>522</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>182</td>
<td>239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65% “major improvement” in control group</td>
<td>1416</td>
<td>1882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70%</td>
<td>151</td>
<td>197</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>82</td>
<td>106</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and power=80%, 1605 patients would be required in each group. If power=90%, the number of patients required in each group would be 2125. Table 1 provides sample size estimates for varying effect sizes (rounded to the nearest 5%) and powers of 80% or 90% with α=.05 for both the observed data (approximately 50% of patients demonstrating major improvement) and the upper and lower bounds of the 95% confidence interval.

Thirteen of 21 patients (62%; 95% confidence interval, 41% to 83%) with MCA strokes who had an initial NIHSS score of at least 4 points improved by at least 4 points. Approximate sample sizes are given in Table 2.

As indicated, these calculations are based on the assumption that a 4-point change in NIHSS score is always clinically significant. However, as indicated above, this assumption may be incorrect. As clinical experience with stroke outcome measures increases, interpretation of the results of clinical trials will gain increasing sophistication.

TABLE 2. Sample Sizes for Patients With MCA Stroke With Initial NIHSS Score ≥4 (α=.05)

<table>
<thead>
<tr>
<th>Response in Treatment Group</th>
<th>n (Per Group)</th>
<th>Power 80%</th>
<th>n (Per Group)</th>
<th>Power 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>60% “major improvement” in control group</td>
<td>1511</td>
<td>2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65%</td>
<td>165</td>
<td>216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80%</td>
<td>91</td>
<td>118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40% “major improvement” in control group</td>
<td>1573</td>
<td>2093</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45%</td>
<td>407</td>
<td>538</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>186</td>
<td>244</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80% “major improvement” in control group</td>
<td>219</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>88</td>
<td>113</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Factors Associated With Failure of Aspirin Treatment

To the Editor:

We read with great interest the article by Bornstein and colleagues suggesting that stroke patients taking lower doses of aspirin are at greater risk of recurrent stroke than those taking higher doses. We would like, however, to raise three issues concerning the data analysis.

1. The study was designed as a matched case-control study but the matching seems to have been ignored in the analysis. An unmatched analysis of a matched design may result in biased estimates of association.

2. Although the formal statistical test for trend (in which the matching was ignored) yielded a value of P=.06 for the relation of aspirin dose with recurrent stroke, the actual unmatched relative risk estimates provide little evidence for a dose-response relation. Using the data in Table 2, we computed the unmatched odds ratios for recurrent stroke associated with doses of 325 mg/d, 250 mg/d, and 100 mg/d, compared with a dose of 500 mg/d, to be 2.1, 1.7, and 1.7, respectively. As the authors suggested, it may be that doses below 500 mg/d do not offer any benefit.

3. As is evident in Table 3, patients in whom aspirin failed to prevent stroke were more likely to have had ischemic heart disease than their matched controls. The authors should have controlled for that variable (through conditional logistic regression) when assessing the association of aspirin dose with stroke recurrence.

Eyal Shahar, MD, MPH
Paul G. McGovern, PhD
J. Michael Sprafka, PhD
Division of Epidemiology
School of Public Health
University of Minnesota
Minneapolis, Minn
Sample size calculations for clinical trials using the NIH Stroke Scale.

L B Goldstein

*Stroke*. 1994;25:1700-1701
doi: 10.1161/01.STR.25.8.1700

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1994 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/25/8/1700.citation