Sample Size Estimates for Clinical Trials of Vasospasm in Subarachnoid Hemorrhage

Kurt T. Kreiter, PhD; Stephan A. Mayer, MD; George Howard, PhD; Volker Knappertz, MD; Don Ilodigwe, PhD; Michael A. Sloan, MD; R. Loch Macdonald, MD, PhD

Background and Purpose—Clinical trials for prevention of vasospasm after aneurysmal subarachnoid hemorrhage (SAH) seldom have improved overall outcome; one reason may be inadequate sample size. We used data from the tirilizad trials and the Columbia University subarachnoid hemorrhage outcomes project to estimate sample sizes for clinical trials for reduction of vasospasm after SAH, assuming trials must show effect on 90-day patient-centered outcome.

Methods—Sample size calculations were based on different definitions of vasospasm, enrichment strategies, sensitivity of short- and long-term outcome instruments for reflecting vasospasm-related morbidity, different event rates of vasospasm, calculation of effect size of vasospasm on outcome instruments, and different treatment effect sizes. Sensitivity analysis was performed for variable event rates of vasospasm for a given treatment effect size. Sample size tables were constructed for different rates of vasospasm and outcome instruments for a given treatment effect size.

Results—Vasospasm occurred in 12% to 30% of patients. Symptomatic deterioration and infarction from vasospasm exhibited the strongest relationship to mortality and morbidity after SAH. Enriching for vasospasm by selection of patients with thick SAH slightly decreased sample sizes. Assuming $\beta=0.80$, $\alpha=0.05$ (2-tailed) and treatment effect size of 50%, total sample size exceeds 5000 patients to demonstrate efficacy on 3-month patient-centered outcome (modified Rankin Scale).

Conclusions—Clinical trials targeting vasospasm and using traditional patient-centered outcome require very high sample sizes and will therefore be costly, time-consuming, and impractical. This will hinder development of new treatment strategies. (Stroke. 2009;40:2362-2367.)

Key Words: cerebral infarction ■ clinical trial ■ subarachnoid hemorrhage ■ vasospasm

How many patients with aneurysmal subarachnoid hemorrhage (SAH) need to be entered into a clinical trial testing a treatment for vasospasm? Typically, sample size calculations include the definition of trial end point, the event rate of the end point, the effect size of the intervention, type 1 ($\alpha$) and type 2 ($\beta$) error rates, and the mean and standard deviation of trial end point in the treated and control groups. The treatment effect size is the key issue when modeling vasospasm studies because the treatment usually affects vasospasm primarily and only secondarily alters the downstream patient-centered clinical outcome. Because other factors affect outcome, the effect size of a treatment on vasospasm does not translate into the same effect size on the clinical outcome. Another consideration is the definition of vasospasm, which ideally is the one with the highest correlation with clinical outcome, although event rates also need to be taken into account.

Here we derive sample sizes for clinical trials for the reduction in vasospasm after SAH based on the premises that any treatment for vasospasm after SAH must be demonstrable on 3-month patient-centered outcome and must show significant added benefit when combined to the standard of care.

Methods

Patient Populations
The Columbia University SAH outcomes project (SHOP) included 540 consecutive patients with SAH admitted between July 1996 and December 2001. The study was approved by the hospital Institutional Review Board, and informed consent was obtained from the patient or a surrogate. SAH was diagnosed by computed tomography (CT) or xanthochromia of the cerebrospinal fluid if the CT did not show SAH. Exclusion criteria were SAH from trauma or rupture of an arteriovenous malformation, admission >14 days after onset (68% were admitted within 1 day of SAH and 95% within 7 days), and age <18 years. To more closely match a putative phase 3 clinical trial population, the current analysis included 342 patients who were...
admitted within 72 hours of SAH and who underwent either surgical clipping or endovascular embolization (Figure 1). Description of the SHOP patients is published elsewhere.1

The tirilizad data included 3567 patients with aneurysmal SAH entered into 4 randomized, double-blind, placebo-controlled clinical trials of tirilazad between 1991 and 1997.2–5 Patients were entered between 1991 and 1997 at 162 centers from 21 countries in Europe, Australia, North America, and Africa. Detailed descriptions are published elsewhere.6–11 Inclusion criteria were ≥18 years of age and SAH on computed tomographic (CT) scan or lumbar puncture secondary to angiographically-confirmed saccular aneurism. Exclusion criteria were severe medical, neurological, cardiovascular, or psychiatric illness and pregnancy or lactating females. Almost all patients received nimodipine. 63% underwent some form of prophylactic hemodynamic management, and 92% had neurosurgical clipping of the ruptured aneurysm. Fifty-five patients (1.5%) underwent endovascular aneurysm treatment, and all were entered in one study.4

Variables used were symptomatic vasospasm and cerebral infarction. Symptomatic vasospasm was defined as a ≥2-point decrease in Glasgow coma score or a 2-point increase in the motor score of the National Institutes of Health Stroke Scale6 lasting for at least 8 hours. Other causes of deterioration, including but not limited to electrolyte imbalance, hydrocephalus, postoperative brain swelling or hemorrhage, seizures, had to be excluded. Confirmation with transcranial Doppler ultrasound or cerebral angiography was recommended but not mandatory. Outcome was assessed 3 months post-SAH using the 5-point Glasgow outcome scale (GOS).9 Calculations were based on the placebo group (n=1378).

Statistical Analysis

The first analysis used the SHOP data. Vasospasm was defined 3 ways. Symptomatic vasospasm was defined as clinical symptoms 5 to 12 days after SAH (worsening headache, neck stiffness, decreased level of consciousness, or focal deficits) with no other clinical or radiological etiology. A stricter clinical definition of vasospasm was delayed cerebral ischemia (DCI): (1) clinical deterioration or (2) infarction on CT scan that was not present on admission or postaneurysm securing procedure CT scan and (3) exclusion of other causes of clinical deterioration.1 Angiographic confirmation was not required. Infarction from vasospasm was defined as CT hypodensity appearing on a CT scan 3 or more days after SAH and not present on a CT before this time or immediately after the aneurysm-securing procedure. Definitions of vasospasm were evaluated for univariate associations (Spearman rho) with 3-month outcome measures. These were mean modified Rankin score (mRS),10 sickness impact profile physical (SIPP), SIP psychosocial (SIIPPS) and SIP total (SIPT) score,11 Glasgow outcome score (GOS), and Lawton instrumental activities of daily living (IADL) score.12 The best definition of vasospasm was taken as the definition with the highest and greatest number of significant associations with outcomes. This definition was used for subsequent analyses.

The same procedure was done in the tirilazad data, for which there were 2 definitions of vasospasm; symptomatic13 and cerebral infarction attributable to vasospasm. The only outcome measure was GOS at 3 months. All were based on investigator opinion without central review.

The percent of patients developing vasospasm (the event rate) for different definitions of vasospasm were determined in each database. Groups enriched for patients fitting the different definitions, which were DCI and cerebral infarction, were examined based on variables associated with these factors in each database. For SHOP, the groups were all patients (no enrichment), Hunt and Hess grade 2 to 5, 1 to 4 or 2 to 4, Fisher group 2 to 4, Fisher group 3, World Federation of Neurosurgical Societies (WFNS) grade 3 or 4, anterior communicating artery aneurysm location, thick SAH in the anterior interhemispheric fissure, intracerebral hematoma, cerebral edema on admission CT scan, and angiographic vasospasm on admission.13–15 For tirilazad data, the groups were based on age, history of hypertension, WFNS grades 2 to 4, thick SAH on admission CT, and intraventricular hemorrhage.7,16

Enriched populations then were evaluated for maximal differentiation of clinical outcomes between patients with and without DCI or cerebral infarction. Variance (σ² where σ is the standard deviation of the mean mRS score for the enrichment group of interest), δ (standardized difference in group means=|mean mRS with DCI or cerebral infarction-mean mRS without DCI or cerebral infarction|/σ) and plots of correlations between DCI or cerebral infarction and clinical outcomes at 3 months were examined. Correlations were calculated using Spearman rho and Pearson χ² statistics. The most sensitive clinical outcome for detecting adverse effects of vasospasm was taken as that with the best combination of low variance, high δ, and high correlation of outcome between groups with and without vasospasm.

Treatment Δ was calculated as the rate of DCI or cerebral infarction × β/2 and sample size using means of the clinical outcome calculated using:

\[ N_e = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \times (\delta × \left(\frac{1 + \lambda}{\lambda}\right))}{(M_u - M_d)^2} \]

where \( N_e \) is the number of patients in the control group, \( Z_{1-\alpha/2} \) is the value on the abscissa of a normal distribution with mean=0 and standard deviation=1 for the specified α, \( Z_{1-\beta} \) is the value on the abscissa of a normal distribution with mean=0 and standard deviation=1 for the specified β, δ is the variance in outcome, A is the ratio of group sizes, and \( M_u - M_d \) is the difference in clinical outcome true means between groups.17 We used α=0.05, β=0.80, and assumed 2-sided testing and equal group sizes. Sample sizes using the dichotomous mRS or GOS were calculated using:

\[ n = \frac{\left( \sqrt{Z_{\alpha}^2 + Z_{\beta}^2 + P_{CQ}^2} \right)}{\Delta_{\lambda}} \]

This is the χ² approximation for binary outcome measures assuming equal group sizes with >15 per group, uniform allocation and equal group sizes (λ=1). Again, α=0.05 and β=0.8 and PC=event rate (as a proportion) for outcome of interest (mRS or GOS) in the control.
group, PT=event rate (as a proportion) for outcome of interest in the treated group, ΔA=PC-PT and \( \hat{P} = \frac{(P_c + \lambda P_0)}{1 + \lambda} \) = weighted average of 2 event rates and Δ\( \hat{Q} = 1 - \hat{P} \).

Effect size of DCI or cerebral infarction on clinical outcome (mean scores and standard deviations as well as proportions with favorable clinical outcome with or without DCI or cerebral infarction) was calculated as:

Effect on clinical outcome =[(event rate of DCI)×(clinical outcome with DCI)]+(1-event rate)×(clinical outcome without DCI)]÷[(1-(event rate)×(1-effect size on DCI))]×(clinical outcome without DCI).

For each enrichment group, we calculated the mean difference and standard deviation in clinical outcome for patients with and without vasospasm for each of the clinical outcomes. We calculated standardized effect sizes on 3-month clinical outcome that would result from 25%, 50%, and 75% vasospasm treatment effect sizes in a putative treated group compared to the placebo group. Sample size tables were constructed for the 3 treatment effect sizes across a range of DCI rates, variances, and δ values.

Results

Definition of Vasospasm

In the SHOP data, DCI exhibited the strongest correlations with outcomes (Figure 2) and was superior to symptomatic vasospasm and infarction from vasospasm in 9/11 outcome instruments. Infarction from vasospasm was superior to symptomatic vasospasm and DCI for the 3-month TICS score (1/11 outcomes), although the correlations using symptomatic vasospasm were superior only for the 3-month SIPP score (1/11 outcomes). Correlations of DCI with outcomes were highest with the mRS, GOS, and TICS. Variance and δ for 3-month mRS scores in patients with and without DCI also were better than for other clinical outcomes. In the tirilazad data, cerebral infarction gave higher correlations (Spearman rho=0.370 for dichotomous or 0.418 for 5-point GOS) with GOS than symptomatic vasospasm (Spearman rho=0.140 for dichotomous or 0.176 for 5-point GOS). All correlations were significant (P<0.00001). Delta was higher for cerebral infarction (0.467) compared to symptomatic vasospasm (0.145). Therefore, DCI and cerebral infarction were used for subsequent calculations.

Enrichment Strategies

Enrichment for patients with DCI in the SHOP data led to rates of DCI as high as 41% compared to 20% for the whole population (Table 1). The highest rates of enrichment excluded more than 50% of patients. Subsequent analyses included only enrichment groups that included more than 50% of patients. Enrichment did increase event rates, however the differences in outcome scores were essentially unchanged while the variance increased (Table 2). Only the Fisher 2 to 4 group yielded significantly reduced sample size requirements compared to the SAH population as a whole or to other enrichment groups. In the tirilazad data, enrichment increased cerebral infarction from 27% to a maximum of 30% (Table 1). Enrichment strategies had minimal effects with only about 10% differences in sample sizes and no group with a combination of lowest variance and highest δ and correlation with outcome.

Effect Size of Treatment

Effect sizes for treatment of vasospasm can be estimated from the fasudil, nimodipine, and CONSCIOUS-1 studies. Nimodipine decreased cerebral infarction from 33% to 22% (relative risk reduction 33%). Fasudil reduced angiographic vasospasm from 53% to 40% (relative reduction of 26%), symptomatic vasospasm from 42% to 35% (17%), and infarction from vasospasm from 35% to 16% (54%). Clazosentan, 15 mg/h, had a 66% relative risk reduction of angiographic vasospasm. We selected effect sizes of treatment on vasospasm of 25%, 50%, and 75%.

Sensitivity Analysis and Sample Sizes

Sample sizes per group over a range of DCI/cerebral infarction rates, δ values (for mean clinical outcomes) and treat-
ment effect sizes based on 90-day outcome data from the
SHOP and tirilazad datasets were calculated (Tables 3 and 4).

Table 3 shows sample sizes per group estimated from SHOP
and tirilazad data for a treatment that decreases the vaso-
spasm end point 25% to 75%. These data use the mean mRS,
but calculations using proportions of patients achieving good
outcome (mRS 0 to 2 or good outcome and moderate
disability on the GOS) were very similar (data not shown).

Table 4 shows the effect on sample sizes per group of varying
the DCI/cerebral infarction event rate and the mean differ-

Table 1. Effect of Enrichment Strategies on Rate of DCI in SHOP and Symptomatic Vasospasm in Tirilazad
Databases*

<table>
<thead>
<tr>
<th>Enrichment Group</th>
<th>n</th>
<th>% of Total SAH Population</th>
<th>DCI/Cerebral Infarction (n, %)</th>
<th>% of Total With DCI/Cerebral Infarction</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOP database</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All SAH</td>
<td>540</td>
<td>100</td>
<td>110 (20)</td>
<td>20</td>
</tr>
<tr>
<td>WFNS grade 3 or 4</td>
<td>114</td>
<td>21</td>
<td>43 (38)</td>
<td>8</td>
</tr>
</tbody>
</table>
| Anterior communi
cating artery aneu
rysms           | 114 | 21 | 36 (32) | 7 |
| Thick anterior in
terhemispheric fissure SAH | 67 | 12 | 21 (31) | 4 |
| Intracerebral hemorrhage | 86 | 16 | 26 (30) | 5 |
| Cerebral edema on admission | 50 | 9 | 8 (16) | 1 |
| Angiographic vasospasm on admission | 61 | 11 | 25 (41) | 5 |
| Hunt and Hess grade 2–5 | 389 | 72 | 93 (24) | 17 |
| 1–4              | 473 | 88 | 97 (21) | 18 |
| 2–4              | 322 | 60 | 80 (25) | 15 |
| Fisher group 2–4 | 434 | 80 | 97 (22) | 18 |
| Fisher group 3   | 165 | 31 | 47 (29) | 15 |
| Tirilazad Database |     |     | 365 (27) | 27 |
| All SAH          | 1378 | 100 | 365 (27) | 27 |
| Age 40–69        | 996  | 72  | 281 (28) | 20 |
| History of hyperten
sion               | 465  | 34  | 141 (30) | 10 |
| WFNS grades 2–4  | 700  | 51  | 204 (29) | 15 |
| Thick SAH on admission CT | 917 | 67 | 260 (28) | 19 |
| Intraventricular hemorrhage | 611 | 44 | 185 (30) | 14 |

*SHOP data includes patients who were not treated as well as those admitted >72 hours after SAH. DCI indicates delayed cerebral ischemia; SAH, subarachnoid hemorrhage; WFNS, World Federation of Neurosurgical Societies.

Table 2. Rates of DCI in a Different Cohorts of SAH Patients in the SHOP Data and Cerebral Infarction in the Tirilazad Data*

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>% of Target Population</th>
<th>Rate of DCI (%)</th>
<th>Variance</th>
<th>Treatment ∆ (50% Treatment Effect Size per Group)</th>
<th>Sample Size</th>
<th>Spearman Rho</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOP data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>540</td>
<td>20</td>
<td>0.965</td>
<td>4.0</td>
<td>0.1013</td>
<td>6116</td>
<td></td>
</tr>
<tr>
<td>All treated</td>
<td>342</td>
<td>100</td>
<td>23</td>
<td>1.035</td>
<td>3.46</td>
<td>1.164</td>
<td>4006</td>
</tr>
<tr>
<td>Hunt and Hess 2–5</td>
<td>257</td>
<td>75.1</td>
<td>25</td>
<td>0.715</td>
<td>3.69</td>
<td>0.0876</td>
<td>7551</td>
</tr>
<tr>
<td>Hunt and Hess 1–4</td>
<td>318</td>
<td>93.0</td>
<td>21</td>
<td>0.914</td>
<td>3.24</td>
<td>0.0960</td>
<td>5522</td>
</tr>
<tr>
<td>Hunt and Hess 2–4</td>
<td>232</td>
<td>67.8</td>
<td>23</td>
<td>0.574</td>
<td>3.50</td>
<td>0.0650</td>
<td>13 059</td>
</tr>
<tr>
<td>Fisher 2–4</td>
<td>306</td>
<td>89.5</td>
<td>24</td>
<td>1.050</td>
<td>3.57</td>
<td>0.1239</td>
<td>3651</td>
</tr>
<tr>
<td>Fisher 3</td>
<td>125</td>
<td>36.1</td>
<td>31</td>
<td>0.609</td>
<td>3.31</td>
<td>0.1026</td>
<td>4938</td>
</tr>
<tr>
<td>Tirilazad data</td>
<td></td>
<td></td>
<td>365 (27)</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All SAH</td>
<td>1378</td>
<td>100</td>
<td>365 (27)</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 40–69</td>
<td>996</td>
<td>72</td>
<td>281 (28)</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WFNS grades 2–4</td>
<td>700</td>
<td>51</td>
<td>204 (29)</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thick SAH on admission CT</td>
<td>917</td>
<td>67</td>
<td>260 (28)</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Excludes patients not treated or admitted >72 hours after SAH in SHOP. Outcome is mRS for SHOP data and GOS for tirilazad data. Treatment ∆ is the true difference in group mean scores on the mRS between patients in the hypothetical treatment and placebo groups. CT indicates computed tomography; DCI, delayed cerebral ischemia; SAH, subarachnoid hemorrhage; WFNS, World Federation of Neurosurgical Societies.
ence in outcome between groups. Again, results using dichotomous outcomes were similar (data not shown). These scenarios indicate that efficacy trials will need powerful treatment effects, and large baseline event rates to demonstrate significant effects of a treatment for vasospasm on clinical outcome unless thousands of patients can be randomized.

### Discussion

This analysis shows that clinical trials of treatments directed at decreasing vasospasm require large sample sizes to demonstrate efficacy on 90-day patient-centered outcomes such as the GOS and mRS. These sample sizes would need to be increased further to account for withdrawals and patients lost to follow-up. Although enrichment strategies may decrease sample sizes, they do not reduce the number of patients screened or the length of time to conduct the trial. Given the relatively small size of the target population, such trials may not be scientifically practical or financially possible.

Sample size calculations include definition of trial end point, event rate of the end point, effect size of the intervention, type 1 (α) and type 2 (β) error rates, and mean and standard deviation of trial end point in the treated and control groups. How can these be altered to reduce sample sizes to practical values? We already used relatively lax type 1 (0.05) and 2 (0.80) error rates; more stringent ones would increase sample sizes further. Clinical outcomes like the mRS and GOS are not specific for detecting transient ischemia and may be insensitive to subtle yet meaningful changes in level of functioning. For example, 50% of patients with outcomes on the mRS at the best level exhibit significant cognitive impairment when evaluated with detailed neuropsychological testing. Whether other clinical outcome measures would be more sensitive is unknown. A key feature of this analysis is that effect size on vasospasm translates into very small effect size on mRS or GOS. This is because factors unrelated to vasospasm and DCI contribute to poor outcome and are not affected by treatment of vasospasm. An assumption of this analysis is that treatment affects only vasospasm-related outcome so treatments that have additional beneficial effects may require smaller sample sizes.

Other methods to decrease sample size are to use surrogate outcome measures, in this case DCI/cerebral infarction from vasospasm, or other statistical methods for analyzing outcome on the mRS or GOS such as the sliding dichotomy, ordinal logistic regression, global statistics, and shift analysis.

### Conclusion

Clinical trials of strategies to reduce mortality and morbidity from cerebral vasospasm may require prohibitively large sample sizes. This may be an obstacle to development of new treatments.

### Disclosures

None.

### References


