Targeting Neuroprotection Clinical Trials to Ischemic Stroke Patients With Potential to Benefit From Therapy

Christopher J. Weir, PhD; Markku Kaste, MD; Kennedy R. Lees, MD, FRCP; for the Glycine Antagonist in Neuroprotection (GAIN) International Steering Committee and Investigators

Background and Purpose—Clinical trials of neuroprotective drugs have had limited success. We investigated whether selecting patients according to prognostic features would improve the statistical power of a trial to identify an efficacious treatment.

Methods—Using placebo data from the Glycine Antagonist in Neuroprotection (GAIN) International and National Institute of Neurological Disorders and Stroke (NINDS) recombinant tissue plasminogen activator (rtPA) clinical trials, we developed and validated simple prognostic models for stroke trial end points: Barthel Index, modified Rankin Scale, National Institutes of Health Stroke Scale, and Glasgow Outcome Scale. Using these models, we simulated 1000 clinical trials and estimated, under several hypothetical treatment effect patterns of neuroprotection, the effect on statistical power of including only patients with moderate prognosis. We calculated the number of patients that would have to be enrolled to maintain the statistical power achieved in selecting the whole trial population. Reanalysis of actual data from the NINDS rtPA trials confirmed the results independently.

Results—Selecting patients with moderate prognosis (predicted probability of favorable outcome 0.2 to 0.8) enabled a sample size reduction, without loss of statistical power, of between 54.6% (51.3% to 57.6%) and 68.6% (66.0% to 71.1%), depending on the treatment effect pattern and outcome measure. These benefits were largely due to the exclusion of patients with poor prognosis.

Conclusions—Targeting patients with potential to benefit enables a substantial sample size reduction without compromising statistical power or duration of recruitment. As part of a broader trial design strategy, informed use of prognostic data available acutely would help in identifying effective neuroprotective treatments. (Stroke. 2004;35:2111-2116.)

Key Words: clinical trials • neuroprotection • prognosis

Researchers have yet to identify therapeutic benefit from any neuroprotective drug in acute ischemic stroke. There are several key areas in stroke clinical trial design: choosing the appropriate stroke onset to treatment time window, identifying the optimal dose, targeting the appropriate patient population (given the drug mechanism), and selecting appropriate and timely outcome measures. Failures in these areas, as well as investigations of drugs that were genuinely ineffective, account for the lack of success in neuroprotection.

Selecting patients for ischemic stroke clinical trials on the basis of pathophysiology may reduce the heterogeneity of patients enrolled and hence may improve the statistical power. Increasing the prognostic homogeneity among trial subjects may complement such targeting of salvageable tissue. When the prognosis of a patient is good after stroke, spontaneous recovery occurs frequently. Conversely, patients with very severe stroke may be destined for a poor outcome, regardless of any intervention. Hence, it may be difficult to demonstrate therapeutic benefit of a neuroprotectant agent in subjects at the extremes of the prognostic spectrum; patients with intermediate prognosis may offer the most information on efficacy.

Prognostic markers have been implicit in the entry criteria for trials of neuroprotection and thrombolysis through the exclusion of individuals with a mild neurological deficit and comatose patients. To our knowledge, no trial in acute ischemic stroke has used a prognostic model explicitly to delineate inclusion criteria.

In severe head injury, systematic selection of patients with moderate prognosis allows the sample size of a trial to be reduced by >30% without loss of statistical power.
Prognostic Models

Models for predicting outcome were essential for the selection of patients according to prognosis and in simulating outcomes according to specific treatment effect patterns under neuroprotection. A review of prognostic models found no models predicting the outcomes we sought to investigate with the use of data readily available acutely. We therefore developed and validated simple and generalizable prognostic models. The aim was not to produce definitive models for use in clinical practice but rather to illustrate the patient selection strategy under investigation. We considered outcome measures commonly used to assess favorable outcome at 90 days in stroke trials: National Institutes of Health Stroke Scale (NIHSS) ≤1, Barthel Index (BI) ≥95, modified Rankin Scale (mRS) ≤1, and Glasgow Outcome Scale (GOS)=1. We also combined these in a "global" outcome. A basic prognostic model for each outcome measure was developed from ischemic stroke patients in the placebo group of GAIN International. We externally validated the models using placebo data from the NINDS rtPA trials. The GOS model was developed and validated internally using NINDS rtPA placebo data. All models incorporated age and baseline stroke severity on the NIHSS.

Simulations and Selection Strategies

Baseline clinical data were generated for placebo and active treatment groups by randomly sampling patients with replacement from the placebo group of the source trial. For each combination of outcome measure, treatment effect pattern, and patient selection strategy, the process was repeated to simulate data for 1000 trials. Placebo group outcomes for a given outcome measure were simulated by applying the relevant prognostic model to the clinical data of each patient. Active treatment group outcomes were generated similarly, but a specific treatment effect (Table 1), plausible under neuroprotection, was added to the prognostic model. For example, the reduced effect in lacunar stroke would be relevant in investigating a neuroprotectant that showed little preclinical evidence of protection of white matter or subcortical neurons.

Statistical power under the global outcome depends on the correlations among its component scales; the correlations must therefore be preserved in simulating outcomes. The probability of favorable outcome for each patient was calculated separately for each outcome measure with the use of the prognostic models. Inter correlate among outcome measures were estimated from NINDS rtPA data. Global outcome data were then simulated.

We simulated groups of 600 and 1400 ischemic stroke patients (the approximate sizes of the NINDS rtPA trials and GAIN International, respectively). We applied 2 selection strategies to these groups: all patients (equivalent to using the entry criteria of the source trial) and patients with moderate prognosis (probability of favorable outcome between 0.2 and 0.8). We then explored whether selecting patients with moderate prognosis provided more information about treatment efficacy.

<table>
<thead>
<tr>
<th>Treatment Effect Pattern</th>
<th>OR (Favorable Outcome)</th>
<th>Proportion Favorable*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant† [1]</td>
<td>1.32</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>[2]</td>
<td>1.54</td>
</tr>
<tr>
<td>Increased effect when prognosis moderate‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of good outcome &lt;0.2</td>
<td>1.10</td>
<td>0.095</td>
</tr>
<tr>
<td>Probability of good outcome 0.2 to 0.4</td>
<td>1.32</td>
<td>0.295</td>
</tr>
<tr>
<td>Probability of good outcome 0.4 to 0.6</td>
<td>1.82</td>
<td>0.499</td>
</tr>
<tr>
<td>Probability of good outcome 0.6 to 0.8</td>
<td>1.32</td>
<td>0.687</td>
</tr>
<tr>
<td>Probability of good outcome &gt;0.8</td>
<td>1.10</td>
<td>0.845</td>
</tr>
<tr>
<td>Reduced effect in lacunar stroke‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacunar stroke</td>
<td>1.11</td>
<td>0.600</td>
</tr>
<tr>
<td>Nonlacunar stroke</td>
<td>1.43</td>
<td>0.273</td>
</tr>
<tr>
<td>Increased effect when limb weakness present‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No limb weakness</td>
<td>1.11</td>
<td>0.600</td>
</tr>
<tr>
<td>Limb weakness present</td>
<td>1.35</td>
<td>0.313</td>
</tr>
</tbody>
</table>

*The event rate on the BI resulting from the odds ratio (OR) if placebo event rates were as per GAIN International.
†OR [1] and its upper 95% CI limit [2] from the latest Cochrane Review of intravenous tPA efficacy.
‡For nonconstant treatment effect patterns, the overall OR equals 1.32 (constant effect [1]). The exception is the prognosis-dependent effect (overall OR, 1.28).
For each batch of simulated trials, the power was calculated as the proportion of trials that identified a statistically significant treatment benefit. We estimated the number of patients with moderate prognosis that would have to be enrolled to maintain the statistical power achieved by retaining the whole trial group. The selection strategy was also judged on the proportion of patients remaining eligible because this would affect the duration of trial recruitment.

We evaluated each selection strategy under a range of trial characteristics. This reflects the uncertainty regarding the likely treatment effect pattern. We sought to identify which strategy gave the greatest statistical power and which would be least sensitive to changes in treatment effect pattern. Finally, we applied the prognostic selection strategy to data from part II of the NINDS rtPA trials and analyzed the data as in the original trial protocol to assess its impact on trial size and conclusions.

### Statistical Analysis

Simulated trials were analyzed by unadjusted χ² test (individual end points) and generalized estimating equations (global end point). Standard theory determined the sample size required to achieve a given statistical power for a binary outcome. Simulations were programmed with the use of SAS (version 8.2; SAS Institute Inc).

### Results

#### Prognostic Models

We used data from 738 ischemic stroke placebo group patients (GAIN International) and 312 placebo group patients (NINDS rtPA trials). The prognostic models for the BI, mRS, and NIHSS outcomes, developed with the use of NINDS data, also fitted adequately (Hosmer-Lemeshow probability values all >0.05). When applied to NINDS trial patients, the proportions of outcomes correctly predicted were 70.5%, 76.0%, and 79.5% for the BI, mRS, and NIHSS outcomes, respectively. The prognostic model for the GOS, developed with the use of NINDS data, also fitted adequately (P=0.81), with a cross-validated predictive accuracy of 73.7%.

#### Selection Strategies

Table 2 shows the power achieved for trials based on a pool of 1400 patients simulated from GAIN International data and compares the strategy of selecting patients according to prognosis with the inclusion of all patients eligible for GAIN International. Choosing patients with moderate prognosis invariably increased trial efficiency. Selecting patients on the basis of a predicted probability of favorable outcome between 0.2 and 0.8 generally maintained or increased the statistical power obtained by including all subjects, despite studying substantially fewer patients. For example, for a uniform treatment effect and the NIHSS outcome, power increased from 56.8% to 67.8% despite a reduction in the sample size from 1400 to 806.

Selecting subjects according to prognosis enables substantial sample size reduction (Table 3), regardless of the treatment effect pattern or outcome measure. The smallest reduction was 54.6% (95% CI, 51.3% to 57.6%) when the treatment effect favored nonlacunar strokes and outcome was measured on the mRS. The greatest reduction was 68.6% (95% CI, 66.0% to 71.1%) for the BI outcome when the treatment effect was dependent on the prognosis. Together with the proportion of patients who would be selected, the required sample size enables us to assess whether a selection strategy would require a longer duration of recruitment.

### Table 3. Sample Size Required Compared With a Conventional Trial

<table>
<thead>
<tr>
<th>End Point</th>
<th>Patients Included</th>
<th>Actual Sample Size Selected</th>
<th>Uniform Treatment Effect</th>
<th>Moderate Prognosis Do Better</th>
<th>Smaller Treatment Effect in Lacunar Stroke</th>
<th>Increased Effect When Limb Weakness Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
<td>1400</td>
</tr>
<tr>
<td>BI Moderate prognosis (20–80%)</td>
<td>812</td>
<td>584 (540 to 628)</td>
<td>440 (404 to 476)</td>
<td>562 (524 to 602)</td>
<td>554 (516 to 592)</td>
<td></td>
</tr>
<tr>
<td>mRS Moderate prognosis (20–80%)</td>
<td>834</td>
<td>632 (586 to 678)</td>
<td>470 (436 to 508)</td>
<td>636 (594 to 682)</td>
<td>566 (528 to 604)</td>
<td></td>
</tr>
<tr>
<td>NIHSS Moderate prognosis (20–80%)</td>
<td>806</td>
<td>592 (548 to 636)</td>
<td>498 (460 to 536)</td>
<td>576 (538 to 618)</td>
<td>524 (488 to 560)</td>
<td></td>
</tr>
</tbody>
</table>

For each end point, the sample size (95% CI) required to maintain the same statistical power as the strategy of including all 1400 patients is shown.
When patients with moderate prognosis are selected, the recruitment duration would be shorter than when all patients are included, since the required sample size is lower than the number of patients selected from the original pool.

**Evaluation on Independent Data**

We assessed the validity of the selection strategy in a different patient population using NINDS trial data. In trials of 600 patients, the larger constant treatment effect (odds ratio 1.54) was simulated, while nonconstant effects had larger differences between subgroups than those shown in Table 1. Comparing results of trials simulated from NINDS data with Table 2 showed the same pattern of efficiency changes. For the moderate prognosis selection strategy, the required sample size results for the NINDS data confirmed a consistent benefit. The NINDS data also allowed investigation of the GOS and global outcome measures (Table 4), for which the moderate prognosis selection strategy enabled consistent sample size reductions across treatment effect patterns.

When we applied the prognostic selection strategy to NINDS (part II) data, between 161 and 196 of 333 patients remained eligible, depending on the outcome measure. The trial findings remained statistically significant (for example, global outcome $P=0.027$, odds ratio for favorable outcome $=1.8$ [1.1 to 3.0]). It became clear that only 1% to 2% of patients were being excluded because of very favorable prognosis. The benefit of selecting patients with moderate prognosis, compared with current trial recruitment strategies, is therefore largely due to the exclusion of those with very poor prognosis. Reanalysis of the NINDS data confirmed this, showing that the trial would remain conclusive if we excluded only patients with poor prognosis (Figure 1).

**Discussion**

Our results demonstrate that basing entry criteria on a simple prognostic model would enhance the statistical power of trials of neuroprotectant agents, allowing reduced sample sizes or the detection of more modest treatment effects for a given trial size. The benefit is due to more patients having the potential, under efficacious treatment, to cross the threshold defining favorable outcome.

Why might patient selection based on prognosis be advantageous? By considering the effects of clinical features in combination, rather than setting a separate threshold for each factor, the pool of trial patients contains a broad range of clinical presentations while retaining prognostic homogeneity. For example, consider a conventional trial for patients of all ages that required a baseline NIHSS score of $\geq 4$ as an entry criterion. A woman aged 80 years with an NIHSS score of 3 would thus be excluded. The simple prognostic model places her in the dark gray region in Figure 2: there is sufficient uncertainty about her outcome that she would be informative with regard to treatment efficacy. Conversely, a 70-year-old man scoring 18 on the NIHSS would be included under conventional criteria, despite his having little chance of a favorable outcome.

The prognosis-based strategy was consistently effective across all outcome measures and treatment effect patterns. Such a characteristic is important since we have little prior knowledge of the size or pattern of treatment effect that might be observed. Applying the selection strategy to independent data from part II of the NINDS rtPA trials illustrated its external validity and highlighted that the increased statistical power was largely due to the exclusion of patients with poor prognosis.

A selection strategy that enabled a reduced sample size but required a greatly increased enrollment period for a trial would be unattractive. Increasing the homogeneity of patients

**TABLE 4. Sample Size Required: Evaluation on Independent Data**

<table>
<thead>
<tr>
<th>End Point</th>
<th>Patients Included</th>
<th>Actual Sample Size Selected</th>
<th>Uniform Treatment Effect</th>
<th>Moderate Prognosis Do Better</th>
<th>Smaller Treatment Effect in Lacunar Stroke</th>
<th>Increased Effect When Limb Weakness Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>GOS Moderate prognosis (20–80%)</td>
<td>358</td>
<td>384 (354 to 414)</td>
<td>234 (202 to 264)</td>
<td>348 (318 to 380)</td>
<td>408 (366 to 450)</td>
<td></td>
</tr>
<tr>
<td>Global Moderate prognosis (20–80%)</td>
<td>336</td>
<td>351 (330 to 375)</td>
<td>156 (147 to 167)</td>
<td>391 (367 to 418)</td>
<td>384 (359 to 412)</td>
<td></td>
</tr>
</tbody>
</table>

For each end point, the sample size (95% CI) required to maintain the same statistical power as the strategy of including all 600 patients is shown.

![Figure 1. Effect on NINDS (part II) results of excluding patients with poor prognosis. Patients were excluded if the probability of achieving good outcome was less than each threshold. ● indicates log-odds ratio for treatment effect; |, 95% CI for log-odds ratio; ✖, percentage of original trial patients selected; and . . . , no treatment benefit.](http://stroke.ahajournals.org/doi/figure/1)
while sacrificing the generalizability of trial results is also undesirable. The prognosis-based method excluded a moderate proportion of patients (~40%) compared with a conventional trial. It may also reduce recruitment duration while maintaining the power that would be achieved by a conventional trial.

For treatments such as thrombolysis with associated risk, it is important to assess the risk–benefit balance across a broad spectrum of patients to identify all who might potentially benefit. Although this conflicts with choosing patients more selectively, the issue would be less important in evaluating neuroprotective treatments that have little associated risk.

GAIN International included noncomatose adults who were previously independent, had symptoms that included limb weakness, and could be treated within 6 hours of stroke onset. The NINDS rtPA trials recruited ischemic stroke patients who presented with measurable deficit on the NIHSS, had a baseline CT scan free of intracranial hemorrhage, and could be treated within 180 minutes of onset. Both trials excluded severely hypertensive patients. Apart from the short time windows, inclusion criteria were general, and there were no explicit prognostic entry criteria. These data were therefore suitable for the present study and allowed evaluation of selection strategies on neuroprotective and thrombolytic trial populations. Nonetheless, patients who would have been informative about treatment efficacy may have been excluded from the original clinical trials and were unavailable for our study. By extending the pool of available patients, it may be possible to increase the benefits and generalizability of prognosis-based selection. Although we studied outcome measures frequently used in acute stroke trials, our findings relate only to those end points.

Prognosis-based patient selection for trials could result in stepwise regulatory approval that is initially given for a subset of patients and subsequently extended to a wider population. Similarly, memantine has been approved for mild to moderate Alzheimer disease and was subsequently investigated in severe disease. Stepwise approval seems unlikely to be necessary, however, since the proposed approach will not restrict entry on the basis of any single factor; the trial population will still include patients with a wide range of age and baseline severity. Regulatory approval should be based on this wide range.

Implementing restrictive trial entry criteria may diminish the motivation of investigators and may result in the inadvertent omission of eligible patients. If prognosis-based patient selection proved unacceptable to investigators, one could pursue an alternative strategy. After broad entry criteria are used, data analysis could incorporate a prespecified primary analysis on the moderate prognosis subgroup (defined with the use of pretreatment data). Such a design would optimize the investigation of efficacy while maximizing the potential generalizability of the results. The risk of this approach is that regulatory authorities may consider only the intent-to-treat analysis, in which the diluted treatment effect may be neutral.

Targeting patients with appropriate prognosis is only one of several key aspects of clinical trial design in neuroprotection. Time from stroke onset to treatment is a critical patient selection criterion, because the earlier the time of therapeutic intervention, the greater the chance of observing a treatment benefit. Patients should be selected who have a tissue substrate on which the neuroprotective compound may act. Accurate estimation of the dose–response curve from early-phase trials is also essential. Our methodology could also determine the impact on statistical power and sample size of selecting patients according to the presence of potentially salvageable hypoperfused tissue, as defined by diffusion- and perfusion-weighted MRI.

The complexity of designing clinical trials of neuroprotective drugs in acute ischemic stroke presents investigators with a considerable challenge. The wide-ranging recommendations in place reflect this, and no single technique will provide a solution. Nevertheless, as part of a multifaceted strategy, informed use of prognostic data available acutely will help to ensure that identification of an effective neuroprotective treatment remains an achievable goal.

Acknowledgments

This study was supported by a UK Medical Research Council Career Development Fellowship to Dr. Weir. The authors thank the NINDS rtPA Stroke Study Group for provision of the rtPA trial data. GAIN International Steering Committee members are as follows: K. Asplund, A. Carolei, S.M. Davis, H.-C. Diener, M. Kaste, J.-M. Orgogozo, J. Whitehead, K.R. Lees. The GAIN International Trial was sponsored by GlaxoSmithKline (previously GlaxoWellcome).

References


Targeting Neuroprotection Clinical Trials to Ischemic Stroke Patients With Potential to Benefit From Therapy

Christopher J. Weir, Markku Kaste and Kennedy R. Lees
for the Glycine Antagonist in Neuroprotection (GAIN) International Steering Committee and Investigators

*Stroke.* 2004;35:2111-2116; originally published online July 8, 2004;
doi: 10.1161/01.STR.0000136556.34438.b3

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 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/35/9/2111