Neuroprotection Is Unlikely to Be Effective in Humans Using Current Trial Designs

James Grotta, MD

What we need to carry out in the laboratory and at the bedside are experiments. If our experiments have been positive in the laboratory but negative at the bedside, it is logical to me that at the bedside we need to better emulate the conditions under which the laboratory experiment turns out positive; in other words, we need to do the “rat experiment” in man. Those clinical studies that have adhered to this dictum1–3 have been the only positive clinical trials to date.

Figure 1 depicts the general design of the rat transient middle cerebral artery occlusion model that is most often used to test neuroprotective drugs. It also depicts the general design of the sort of clinical trial that I postulate must be done to get positive results in stroke patients. There are 4 main areas where clinical trials have departed most from this model. In rats, we take pains to produce lesions of standardized severity in order to better detect a treatment effect; we start by giving the drug soon after the onset of stroke and then determine how much this time to treatment (TTT) can be lengthened; we use models of temporary rather than permanent arterial occlusion; and we increase doses of drug until we see a therapeutic effect.

These factors must all be addressed in the design of future clinical trials, and they are listed in Figure 2 along with one important point. We need to find treatments that are substantially more potent than those that have failed in clinical trials to date.

Standardize Stroke Severity

There are two reasons why standardizing stroke severity is important in our experiments. Both relate to optimizing the ability to see a treatment effect between the drug and placebo. First, if stroke severity is too great, then animals die whether or not they receive treatment. If severity is too mild, the lesion is so small that any differences cannot be detected. Second, if, by chance, the distribution of initial stroke severity varies between the treatment groups, the effect of this imbalance could be much greater than any effect of the treatment.

Attempts to standardize stroke severity in patients randomized into clinical trials is important for the same reasons. The severity of stroke is reflected in the National Institute of Health Stroke Scale (NIHSS). In clinical studies, the baseline NIHSS is clearly the most important variable predicting outcome.4 Using low and high NIHSS cutoffs and ensuring that treatment groups are matched in distribution of NIHSS scores may be one way to achieve standardized stroke severity in our trials.

Stroke standardization might be better accomplished by assessing tissue viability using MRI. However, MRI criteria for predicting tissue outcome is still uncertain, and adhering to a very narrow time window often does not allow for ancillary tests to determine tissue viability. Keeping the time to treatment brief may itself help standardize stroke severity since all patients would have relatively brief, and therefore more reversible, ischemia.

Other variables affecting stroke severity in rats that have not been controlled in most human studies are the number of vessels occluded and location in cortex or white matter. Only 1 study5 was designed to limit patients enrolled to only one type of vascular lesion. The results were positive, again reflecting the wisdom of designing clinical studies to closely emulate what we do in animals. Noninvasive techniques such as transcranial ultrasound and MRI could help us standardize these variables.

Shorten Time to Treatment

In the laboratory, the investigator always starts with a brief TTT and then gradually prolongs it until an effect is no longer seen. This is just the opposite of what has been done in all clinical studies of neuroprotection.

Important lessons can be learned by comparing laboratory with clinical results using thrombolysis to achieve tissue reperfusion. In the laboratory, reperfusion must occur within 2 to 4 hours to see a reduction of infarct size.6 Clinical studies using intravenous rt-PA begun within 3 hours showed a positive effect,1 with more benefit associated with earlier treatment within that window. If begun after 3 hours, rt-PA had little or no benefit.6 Considering that it takes 30 to 90 minutes for a clot to dissolve after beginning intravenous rt-PA, the TTT for reperfusion in clinical studies correlates very nicely with what was found in laboratory models. Now we can carry out further studies in selected patients to see if we can find benefit with longer TTT.

Preclinical studies have shown that all neuroprotective drugs are less effective the later they are given, and most are ineffective if started more than 2 to 4 hours after the onset of ischemia. Yet no clinical trial has yet included enough patients within that 4-hour time window to reach any conclusions about efficacy. Pharmaceutical companies and their consultants, naturally interested in establishing the largest
market for their drug, have abandoned the laboratory data and extended the TTT in most clinical studies to 6 hours or more.

**Combine Neuroprotection with Reperfusion**

Another lesson from laboratory stroke models is that neuroprotective therapy is generally more effective if given to animals with reversible rather than permanent arterial occlusion. One obvious reason is that for a neuroprotective drug to work, it must reach the injured tissue. Also, neuroprotective drugs are particularly effective by targeting cellular events triggered when injured brain tissue is reperfused.

Is it possible to design a study combining reperfusion and neuroprotection? In a recent safety study of lubeluzole combined with rt-PA,7 all patients received intravenous rt-PA within 3 hours of stroke onset, and were begun on study drug or placebo within 1 hour of starting rt-PA.

**Give Sufficient Dose**

Side effects have been the Achilles heel limiting the doses of neuroprotective drugs given to stroke patients. Recent stroke victims seem particularly vulnerable to cardiovascular or sedative effects of a drug, often not predicted by studies in young rats or in normal volunteers. These considerations make it difficult to predict the blood levels we can expect to reach safely in our acute stroke patients.

In designing a clinical neuroprotective trial, we should begin with a dose escalation study (with blood level correlation) in stroke patients similar to those intended for the pivotal trial. The highest tolerable dose should then be chosen for the pivotal trial. Ideally, the next step would be to demonstrate that the dose chosen has the desired effect on the target biological process.

**Find a More Effective Drug**

Most of the neuroprotective drugs subjected to clinical evaluation have been able to improve outcome by about 50% in well-standardized preclinical stroke models. It is possible that even if we adhere to the principles described in the preceding paragraphs, this effect is not enough to detect in our stroke patients.

Based on our inability to translate preclinical results to the bedside, we probably should no longer move forward with clinical evaluation of a monotherapy that reduces damage by only 50% in a rat. We need to find drugs that reduce damage by 80% in these models. While testing a drug that affects a single pathway in the process of brain injury has scientific and regulatory purity, this strategy has proven ineffective clinically. It is time to look for combinations of drugs that have a stronger effect. We need to work on the regulatory and financial impediments for conducting such trials.

**References**


**Key Words:** animal models ■ clinical trials ■ neuroprotection ■ stroke assessment
Neuroprotection Is Unlikely to Be Effective in Humans Using Current Trial Designs: An Opposing View

Kennedy R. Lees, MD, FRCP

A difficulty with better tolerated drugs that may have a bell-shaped dose-response relationship, such as neutrophil inhibitory factor (ASTIN trial) or BMS-204,352 (POST trials), is that dose selection relying on the maximum tolerated dose may be flawed: instead, some measure of efficacy needs to be incorporated. Here, modern adaptive dose selection designs are going to be profoundly useful. The ASTIN trial uses a design first applied in cancer trials: patient responses are monitored “on line” and used to inform a computer algorithm, which “decides” on the most informative dose to employ at the next randomization opportunity (for discussion, see http://lib.stat.cmu.edu/bayesworkshop/Bayes99.html). This design should revolutionize neuroprotective drug development: not only will the optimal dose be selected, the trial will also decide when enough confidence in the drug is attained to justify progression to phase III; alternatively, it can be used to determine futility with scientific rigor, rather than risking abandoning a potentially useful compound on management whim. Such new designs place exacting demands on the steering and data safety monitoring committees, however, since new skills are required.

Drug and Dose Selection Have Been Flawed

Originally, limited efficacy data from a single laboratory and an acceptable toxicity profile were sufficient to proceed to clinical development. Choice of dose range, the optimal duration of therapy, patient subgroup selection, and the likely extent of benefit were all reserved for testing in clinical trials. The misplaced optimism of such an approach is now recognized, however, and guidance on preclinical testing has been elaborated. Rigorous external scrutiny of preclinical data now occurs in the more informed companies, and clinical development is deferred pending resolution of any deficiencies. This is the first role of the (now mandatory) trial steering committee.

The drug classes under investigation are evolving, and new approaches are still being discovered. Glutamate antagonists have been widely studied. While effective in small animal models, their short time window in the laboratory, failure to protect white matter or oligodendrocytes, and dose-limiting psychoactive effects are notable and perhaps fatal disadvantages. Lubeluzole and eliprodil suffered dose-limiting QTc prolongation. Gavestinel doses were restricted by liver function disturbances. Perhaps the safety and multimodal effects of magnesium, or novel approaches, eg, with proteasome inhibitors, may prevail.

The opinions expressed in this editorial are not necessarily those of the editors or of the American Stroke Association.

From the University Department of Medicine & Therapeutics, Western Infirmary, University of Glasgow, UK.

Kennedy Lees holds grants to investigate stroke trial design, is principal investigator of the IMAGES (sponsored by the UK Medical Research Council) and GAIN International (sponsored by GlaxoWellcome) trials, serves on the data safety monitoring committees for the ASTIN (Pfizer) and DIAS (PAION) trials, and was a member of the STAIR group.

Correspondence to Dr. Kennedy R. Lees, University of Glasgow, University Department of Medicine & Therapeutics, Western Infirmary, Glasgow G11 6NT, UK. E-mail k.r.lees@clinmed.gla.ac.uk

(Stroke. 2002;33:308-309.)

© 2002 American Heart Association, Inc.

Stroke is available at http://www.strokeaha.org

To summarize recent neuroprotective trials, there has been no useful balance of benefit over risk. It is premature to shoot the messenger without evidence either that the results have been misleading or that we have missed trends. Confirmatory trials and/or meta-analysis support the validity of the results, and similar designs have successfully revealed the risks and benefits of thrombolysis. Trial designs have been imperfect, but the message is accurate: the drugs or doses we tried were inadequate. We need safer and better drugs, we need to know when and how to use them, and we need to maximize the efficiency of our trials. All are possible with current designs.

Entry Criteria Have Tightened

Patients may have been poorly selected for trials: the CLASS trials optimistically treated patients up to 12 hours from stroke onset, and the “definitive” lubeluzole phase III trial extended the time window from 6 to 8 hours. At worst, this has limited trial power. Patient selection has been improved for current trials, however. The ASTIN trial has a 6-hour window, but vigorous efforts are keeping the average time to treatment nearer to 4 hours and also discouraging recruitment of patients who narrowly miss exclusion on several criteria. Interactions among entry criteria are increasingly considered. Selection according to a combination of time window and persisting evidence of salvageable tissue (diffusion/perfusion mismatch on MRI) is being used for a sipatrigine trial and in the DIAS trial of desmoteplase. Even in the long-running IMAGES trial with magnesium, in which late treatment is very unlikely to cause harm and so marginal benefits may be worthwhile, incentives are offered to keep the proportion treated early above 40%, and the MR IMAGES substudy selects on the basis of MR mismatch. Higher risk of adverse effects effectively shortens the window of utility: a fixed risk offset against a benefit that diminishes with time translates into overall harm after a finite time delay; in contrast, a completely safe treatment could in theory be given extremely late without offsetting any minimal benefit.
effective: to power a study for a 10% to 12% absolute benefit is unrealistic. Current trials are targeting smaller benefits: GAIN recruited more than 3000 patients between 2 trials; IMAGES seeks a 5.5% improvement in 2700 patients (though even that may be optimistic), but more important, sequential or adaptive sample size designs are now accepted, as evidenced by ASTIN and DIAS. Through stratification, GAIN achieved almost perfect balance in prognostic factors between groups. Methods to limit baseline imbalance in prognostic variables are becoming more sophisticated: adaptive randomization in conjunction with interactive telephone response systems can achieve balance on up to 12 variables without implications for drug supplies and without leading to impractical block sizes. Unfortunately, the most sophisticated systems are still viewed with suspicion and combinations of techniques have yet to be implemented.

Trial endpoints are being better selected: with unselected populations showing a U-shaped distribution of Barthel outcomes, even a marked improvement across the board leads to few patients crossing any arbitrary boundary set in the range of 60 to 90 Barthel points. Thus, recent trials like GAIN have identified trichotomous outcomes, have combined measures on various scales using the global statistical test, or have set variable outcome thresholds according to initial stroke severity. Crucially, through the agency of steering committees, invaluable data from trials and registers involving many thousands of patients such as GAIN are creating a resource that is informing the design of current and future trials using designs such as ASTIN.

Designs Should Change but Must Remain Realistic
Past approaches precluded prior treatment with thrombolysis, yet several putative neuroprotectants will likely be effective only in the presence of reperfusion. Current trials allow thrombolysis according to local guidelines, even where alteplase is used off-label: the POST, IMAGES, and ASTIN trials are examples. Institutional Review Board approval may even be dependent on allowing such treatment. Subgroup analysis of completed trials that allowed thrombolysis has not hinted at improved outcomes with combination treatment. Unfortunately, although protocols encourage early initiation of the putative neuroprotectant, practicalities often dictate that thrombolysis is started or even completed before randomization. These issues still need to be tackled.

Quality of Data Is Paramount
Previously, incentives were offered to accelerate patient recruitment; in recent trials, data quality has taken priority. It is now routine to expect training of staff in trial procedures (NIHSS, SSS, and Barthel), feedback on performance (ASTIN, POST), central reading of images (GAIN, DIAS, POST), and external monitoring—even in academic trials such as IMAGES.

Conclusion
Current trials are applying techniques new to stroke but which have proven worth in other fields. Unfortunately, no trial has yet employed all of the available strategies to maximize its chances of success. Nevertheless, designs are available and in current use that will allow a truly neuroprotective compound to demonstrate its value. We have the technology; we must demand it is properly used.

References

Key Words: clinical trials ■ neuroprotection ■ stroke assessment

Neuroprotection: Establishing Proof of Concept in Human Stroke

Stephen M. Davis, MD, FRACP; Geoffrey A. Donnan, MD, FRACP

The failure to translate the positive effects of a variety of neuroprotective strategies from animal models to human trials has perplexed investigators. This contrasts with the success of tissue plasminogen activator, where efficacy in animal models was translated into positive trial results. Postulated reasons for these expensive trial failures include defective trial design (particularly time window and trial size) and the heterogeneity and complexity of human stroke compared with
animal infarct models. Further explanations might be the lack of entry of neuroprotective compounds into the ischemic brain (suggesting that these agents might require thrombolysis to facilitate access), their lack of benefit in white matter, and the fairly modest effect of neuroprotection in penumbral salvage, even if reperfusion has occurred.

We agree with Grotta that the animal experiments should be replicated as closely as possible in trials, including standardization of stroke severity and appropriately short times to treatment. Importantly, he also mentions that compounds with multiple sites of action on the neurotoxic cascade may be more effective. Lees has pointed out that improved trial design, tightened entry criteria, and more sophisticated endpoint selection have already occurred and should increase the chances of success.

Despite the plethora of negative neuroprotective trials, an array of compounds has been shown to dramatically reduce infarct volume in animal models. It seems quite implausible to us that these experiments should not be translated to human stroke, given a potent agent, adequate drug levels in ischemic tissue, a short time window, lack of toxicity, and well-designed stroke trials. We suggest a stepwise approach, which includes demonstration of a highly significant infarct volume reduction in small animal models (at least 80% as suggested by Grotta), together with functional improvement. A similar benefit should then be shown in larger animal models, perhaps incorporating neuroimaging endpoints as a further intermediate step. Proof of concept studies in humans, using perfusion/diffusion-weighted MRI as a surrogate endpoint, should more reliably indicate the chance of efficacy in phase III trials.

The views of both our debaters clearly indicate that we need to rethink our approach to translational research. We conclude that however sophisticated our current phase III trial designs, the testing of the hypothesis that neuroprotection is an effective stroke treatment requires a far more rigorous approach than was originally perceived.
Neuroprotection Is Unlikely to Be Effective in Humans Using Current Trial Designs
James Grotta

Stroke. 2002;33:306-307

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/33/1/306