Acute Stroke Trials: Strengthening the Underpowered

Anthony J. Furlan, MD

In this issue of Stroke, Muir addresses the looming crisis in acute stroke clinical trial design by illustrating why neuroprotective trials have been seriously underpowered. Unfortunately, this is not a new observation. Samsa and Matchar pointed out 3 statistical reasons neuroprotective stroke trials have been underpowered: (1) Sensitivity of power to small changes in outcome rates. (2) Overestimation of true treatment effect. Typically, neuroprotective stroke trials are powered to detect absolute treatment effects of \geq10\%.

This is likely wishful thinking. Phase III neuroprotective stroke trial sample sizes are usually based on optimistic phase II treatment effects. Furthermore, endpoints vary from trial to trial and may be erroneously selected on the basis of phase II data (eg, citicoline used an unconventional NIHSS analysis and lubeluzole Europe chose mortality).

There is little reason to believe that neuroprotective stroke therapy alone will demonstrate the same magnitude of efficacy as reperfusion stroke therapy with intravenous tissue plasminogen activator under 3 hours (13\% absolute benefit) or intra-arterial thrombolysis at 6 hours (15\% absolute benefit). (3) Underestimation of the minimum clinically important difference. Since stroke is disabling with high long-term care costs, even a very modest treatment benefit on the order of 2\% may result in a net benefit from a population viewpoint.

Cardiology trials have employed this type of analysis to demonstrate the cost-effective benefit of new therapies that are commonly available and relatively safe therapies are beneficial when used indiscriminately in community-based stroke populations, but given the heterogeneity of acute stroke and the risks involved, no neuroprotective or reperfusion therapy is likely to pass that test.

IST and CAST aside, there are 3 notable exceptions to failed acute stroke trials: NINDS, STAT, and PROACT II. There are several likely reasons these trials succeeded while others have failed: (1) Perfusion. All 3 trials tested reperfusion therapies. Neuroprotection efficacy alone without timely reperfusion of ischemic brain may be very difficult to demonstrate. (2) Pathophysiological homogeneity. PROACT II randomized relatively homogeneous stroke patients with a demonstrated stroke etiology (middle cerebral artery occlusion) likely to benefit from the treatment intervention (intra-arterial thrombolysis). By removing some of the "noise" from acute stroke, PROACT II was able to demonstrate treatment efficacy even beyond 3 hours with a small sample size (n=180).

(3) Time. Time is a critical factor on both the near and far ends of the therapeutic window. NINDS and STATS were able to demonstrate treatment efficacy even in nonhomogeneous stroke patients with relatively small sample sizes (n=624 and 500, respectively) because treatment was very early (indeed, the major benefit in NINDS was in patients treated <90 minutes from onset). While PROACT II demonstrated that some patients can be helped after 3 hours, pathophysiological homogeneity probably becomes increasingly critical as time goes on (possible reasons why ECASS II and ATLANTIS failed).

So 2 strategies employed in reperfusion trials to keep sample sizes manageable have been to treat heterogeneous stroke patients very early—which, while desirable, may be...
impractical—or to treat homogeneous stroke patients—which, while desirable, may be time consuming and expensive. The obvious corollary is that reperfusion efficacy will be easiest to prove (or disprove) when homogeneous stroke patients are treated very early. Muir’s model suggests these principles also apply to neuroprotection trials but with the added difficulties of treating a multifactorial process (ischemia) with a single agent and without timely reperfusion. Muir suggests several strategies to increase the proportion of “informative patients” in neuroprotection trials that, while initially more expensive and heretofore unappealing to pharmaceutical marketing divisions, are more likely to result in the first neuroprotection therapeutic breakthrough.

I agree completely with Muir that the problem of stroke heterogeneity has been grossly underestimated in clinical trial design. Is it any wonder clinical efficacy has been difficult to demonstrate with a Rankin scale when we are performing underpowered trials with heterogeneous patients in whom we do not understand how recovery occurs in the first place? Is it not erroneous to lump together infarcts of all shapes, sizes, times, severities, and locations due to various occlusions (or no occlusion or site of occlusion unknown) and trust the statisticians to make sense of it all through randomization into underpowered trials? Not to be overlooked as well, stroke trials of the magnitude this outdated approach requires may be impossible to perform in the absence of an organized, sustained international stroke trial consortium effort analogous to GUSTO or TIMI.

New thinking is urgently needed in stroke clinical trial design if we are to begin to solve the crisis in acute stroke therapy. Muir has added to the growing list of options to the traditional randomized clinical trial, which clearly has failed to produce any stroke therapeutic breakthroughs. Such lessons learned should empower stroke investigators to initiate change at all levels of the drug evaluation process. Otherwise, we will continue to design underpowered trials destined for failure rather than for success. Oh, and by the way, we also need drugs that work in humans.

References

KEY WORDS: controlled clinical trials ▪ neuroprotection