

DIALOGUE

Truths about the NINDS study: setting the record straight

Jeffrey Mann
Salt Lake City, UT
jmannemg@earthlink.net

Competing interests:
None declared

West J Med
2002;176:192-194

Thrombolysis for acute ischemic stroke has been studied for more than a decade, but its efficacy remains controversial. The first study to claim that tissue plasminogen activator (tPA) is effective in the treatment of acute ischemic stroke was a multicenter clinical trial coordinated by the National Institute of Neurological Disorders and Stroke (NINDS) Study Group. The NINDS study's conclusions, published in 1995,¹ were that "treatment with intravenous tPA within 3 hours of the onset of ischemic stroke improved clinical outcome at 3 months . . . [A]s compared with patients given placebo, patients treated with tPA were at least 30% more likely to have minimal or no disability at 3 months."^{1(p1586)} The NINDS study was widely perceived to be a well-executed and analyzed randomized controlled trial, and its results were well received by many medical professionals and the public.

Over the past 5 years, tPA therapy for acute ischemic stroke has entered the mainstream of emergency medical practice in the United States. When the American Heart Association revised its advanced cardiac life support (ACLS) guidelines for the 2000 ACLS handbook, *Guidelines 2000 for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care*, it gave tPA a class I recommendation for the therapy of acute ischemic stroke. The American Heart Association gives a drug a class I recommendation if the evidence in support of its effectiveness is considered homogeneous, consistently positive, and robust. Are the NINDS study's results sufficiently robust to withstand rigorous analysis, and is tPA, therefore, fully deserving of a class I recommendation for the treatment of acute ischemic stroke?

ANALYSIS

Consider the actual presentation of the NINDS study's data in the original *New England Journal of Medicine* article. A careful appraisal of the data presented reveals that the NINDS investigators supplied limited information about the baseline National Institutes of Health Stroke Scale (NIHSS) scores of the patients treated with tPA and those given placebo. The scale is an 11-item clinical evaluation instrument widely used in clinical trials and practice to assess neurologic outcome and degree of recovery; it is scored from 0 to 51, where 51 is maximum disability. Median baseline scores for tPA-treated patients versus those given placebo in part 1 of the NINDS trial were 14 and 14, and in part 2 were 14 and 15. Thus, readers were led to assume that randomization was successful in rendering the two groups equivalent. I made the same as-

Summary points

- To obtain valid results, critical prognostic variables have to be prespecified, and corrected for, in the design of any randomized controlled trial
- Baseline stroke severity is a critical prognostic variable in the use of tissue plasminogen activator (tPA) for acute ischemic stroke trials
- Randomization into the tPA and placebo groups was flawed in the National Institute of Neurological Disorders and Stroke (NINDS) trial
- This flaw in randomization could alone account for the apparent effectiveness of tPA shown in the NINDS trial

sumption but was surprised to discover more than 5 years after the initial publication of the NINDS trial's results that the trial actually had an imbalance in baseline stroke severity randomization. The patients treated from 91 to 180 minutes after stroke onset had far less severe strokes than the control (placebo) group. This important fact came to light with the publication of an article in the December 12, 2000, issue of *Neurology* by Marler and other members of the NINDS tPA stroke study group.²

Table 3 in that article (see our Table) reveals the baseline NIHSS scores by time from stroke onset (in the subgroup of patients treated from 91-180 minutes). Note that more tPA-treated patients had mild strokes (15% difference), and fewer had severe strokes (10% difference). Those differences are significant in absolute terms, but their actual significance becomes apparent when the importance of the following graph (Figure) from the TOAST

Percentage of patients (N = 320) in the 91 to 180-minute subgroups with a specific baseline National Institutes of Health Stroke Scale (NIHSS) score*

Baseline NIHSS score	tPA-treated patients, % (n = 153)	Patients given placebo, % (n = 167)
0-5	19.0	4.2
6-10	24.2	27.5
11-15	17.0	21.0
16-20	21.6	19.8
>20	18.3	27.5

tPA = tissue plasminogen activator
*From Marler et al.²

study (Trial of Org 10172 in Acute Stroke Treatment) is fully appreciated.³ The graph plots the probability of an excellent outcome of an untreated stroke patient against the baseline NIHSS score.

This graph provides crucial information about what would probably happen to an untreated group of patients with acute ischemic stroke. Note the following:

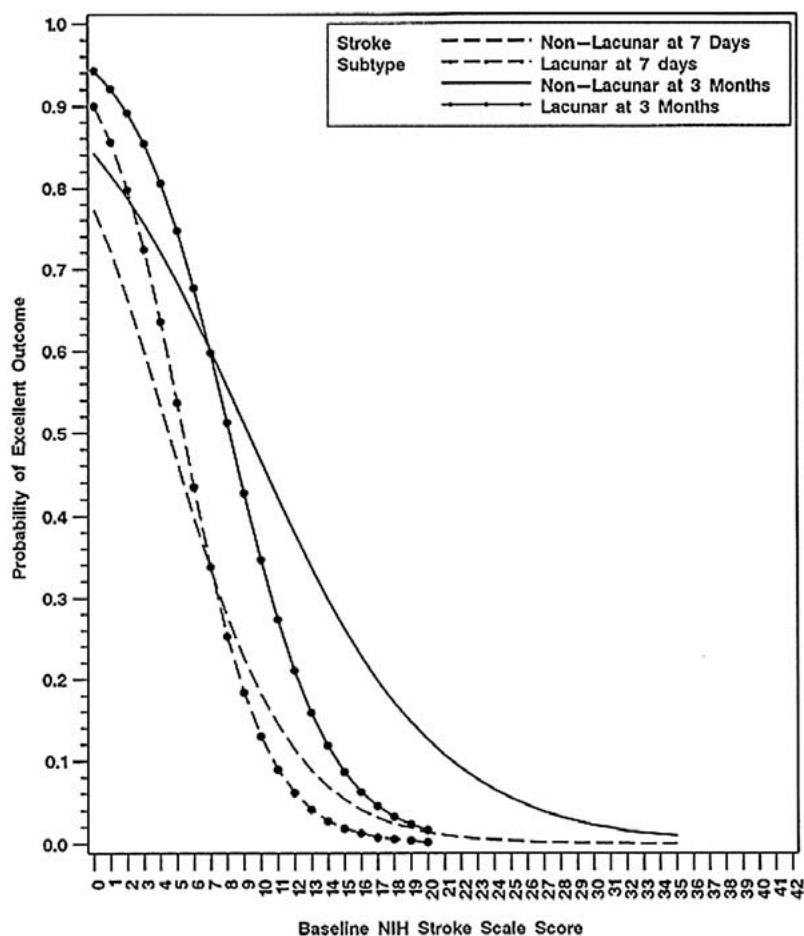
- The curves vary continuously in their slope angle (especially in the higher NIHSS score range), the relationship between the probability of an excellent outcome and baseline NIHSS score is not consistently linear, and the curves have an overall S shape
- The curves are steep at lower baseline NIHSS scores (baseline NIHSS scores between 0 and 15); a small 1- to 2-point absolute difference in baseline NIHSS score in the range of 0 to 15 correlates with a difference in the probability of an excellent outcome of 10% to 20%
- The curves are flat at high baseline NIHSS scores above 20, and stroke patients with a baseline NIHSS score of over 20 have a low probability of an excellent outcome (<10% favorable outcome rate)

On careful consideration of the implications of this graph, it becomes immediately apparent that it is invalid to use statistical techniques based on linear analysis when dealing with an S-shaped curve and that the use of a single median figure for an entire group of patients (with baseline NIHSS scores from 1-37) is a misleading way of representing the baseline stroke severity of the treated patients and those given placebo.

It also becomes immediately apparent that absolute precision is required when randomly allocating patients with lower baseline NIHSS scores (0-15) to treated and placebo groups because a mere 1- or 2-point difference in baseline NIHSS scores between two groups of otherwise equivalent patients who have mild ischemic stroke could result in a 10% to 20% difference in the expected probability of an excellent outcome based on chance alone.

Thus, synthesizing information from both the table and the graph, patients treated with tPA were 15% more likely to have mild strokes than patients given placebo. Such patients had a 75% probability of an excellent outcome based on the natural course of the disease. Patients who received placebo were 10% more likely than tPA-treated patients to have severe strokes. Such patients had a 10% probability of an excellent outcome. Taken together, the skew in randomization itself accounts for the final results of the NINDS trial.

It is well recognized that baseline imbalances in important prognostic variables may culminate in spurious conclusions when not prespecified, and corrected for, in the statistical analysis of a trial's data. This fact has been pre-



Probability of an excellent stroke outcome at 7 days and 3 months as influenced by the baseline score on the National Institutes of Health (NIH) Stroke Scale (from Adams et al³ [Figure 3], with permission)

viously described by Roberts and Torgerson as an important potential bias in the analysis of the results of randomized controlled trials.⁴ The authors note that it is difficult after the fact to correct or adjust for differences in baseline characteristics. Many stroke researchers have suggested that trial investigators should correct for baseline imbalances in stroke severity in the design phase of a trial and that it is better to ensure similarity through a more stringent randomization process or, alternatively, minimize differences that result from chance by doing a much larger study. DeGraba and associates in their article on the performance and evaluation of stroke trials conclude, "It is therefore cautioned that randomization into clinical trials without stratification of stroke severity increases the risk of testing two populations of patients with different clinical courses."^{5(p1211)}

CONCLUSION

The marked imbalance in baseline stroke severity in the 91 to 180-minute groups of the NINDS trial suggests that

the NINDS trial lacks internal validity. Although there was no imbalance in baseline stroke severity in the 0 to 90-minute groups in the NINDS trial, any therapeutic benefit shown by that group has little practical importance because the likelihood of a community physician being able to treat patients who have acute ischemic stroke in less than 90 minutes is very small. That fact suggests that the NINDS trial also lacks external validity, a point emphasized by Hoffman in his analysis of the NINDS trial.⁶ Hoffman stated the following:

An important methodologic concern about the NINDS trial is that it selectively enrolled patients with less than 90 minutes of symptoms. In fact, the study protocol required that investigators recruit equal numbers of “very early” patients (treated in 0-90 min) and “early” patients (91-180 min). The study showed an overall 11%-13% absolute benefit with tPA treatment; however, a recent report by the NINDS authors^[2] clarified that the benefits were greater than this in the “very early” (0-90 min) group, which means that they had to be less than this in the “early” (90-180 min) [*sic*] group. This is extremely important because, in real clinical practice, “very early” patients are almost non-existent. Had a disproportionate sampling mechanism not been built into the trial, it is likely that “very early” patients would have comprised a minimal proportion of those enrolled—and that the apparent benefit of treatment would diminish or even disappear. Thus, if “very early” patients were the group who derived most or all of the benefit in NINDS, the only study to suggest benefit, evidence for use of tPA becomes far more tenuous.

In summary, the recommendations for the use of tPA in patients with acute ischemic stroke were based on an

initial misinterpretation of the results of the NINDS trial and are, therefore, unwarranted. The NINDS investigators may think that tPA works and that no further trials are needed. In fact, Lyden in an editorial in “Controversies in Stroke” wrote, “Perhaps we will find a way to treat patients later than 3 hours, and further studies are needed to push the outer limits of the time window, but within the 3-hour window, no further trials are needed; the drug works. The dictum *primum no nocere* still applies: we must do no harm, either by actively committing an act or by withholding a proven therapy through inaction.”⁷(p2709) The readers of this article should think carefully about these issues and independently decide whether further trials of the use of tPA for acute ischemic stroke are needed.

References

- 1 National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581-1587.
- 2 Marler JR, Tilley BC, Lu M, et al. Early stroke treatment associated with better stroke outcome: the NINDS rt-PA stroke study. *Neurology* 2000;55:1649-1655.
- 3 Adams HP Jr, Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: a report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology* 1999;53:126-131.
- 4 Roberts C, Torgerson DJ. Understanding controlled trials: baseline imbalance in randomised controlled trials. *BMJ* 1999;319:185.
- 5 DeGraba TJ, Hallenbeck JM, Pettigrew KD, Dutka AJ, Kelly BJ. Progression in acute stroke: value of the initial NIH stroke scale score on patient stratification in future trials. *Stroke* 1999;30:1208-1212.
- 6 Hoffman JR. Tissue plasminogen activator for acute ischemic stroke: is the CAEP Position Statement too negative? *Can J Emerg Med CJEM* 2001;3 [electronic journal]. Available at www.caep.ca/004.cjem-jcmul/004-00.cjem/vol-3/2001/v33-183.htm. Accessed February 5, 2002.
- 7 Lyden PD. Further randomized controlled trials of tPA within 3 hours are required—not! *Stroke* 2001;32:2709-2710.