Graphic Reanalysis of the Two NINDS-tPA Trials Confirms Substantial Treatment Benefit

Jeffrey L. Saver, MD; Jeffrey Gornbein, DrPH; Sidney Starkman, MD

Background and Purpose—Multiple statistical analyses of the 2 NINDS-tPA trials have confirmed study findings of benefit of fibrinolytic therapy. A recent graphic analysis departed from best practices in the visual display of quantitative information by failing to take into account the skewed functional importance of NIH Stroke Scale raw scores and by scaling change axes at up to 20 times the range achievable by individual patients.

Methods—Using the publicly available datasets of the 2 NINDS-tPA trials, we generated a variety of figures appropriate to the characteristics of acute stroke trial data.

Results—A diverse array of figures all visually delineated substantial benefits of fibrinolytic therapy, including: bar charts of normalized gain and loss; stacked bar, bar, and matrix plots of clinically relevant ordinal ranks; a time series stacked line plot of continuous scale disability weights; and line plot, bubble chart, and person icon graphs of joint outcome table analysis. The achievable change figure showed substantially greater improvement among tPA than placebo patients, median 66.7% (interquartile range, 0 to 92.0) versus 50.0% (interquartile range, −7.1 to 80.0), P=0.003.

Conclusions—On average, under 3 hour patients treated with tPA recovered two-thirds while placebo patients improved only half of the way toward fully normal. Graphical analyses of the 2 NINDS-tPA trials, when performed according to best practices, is a useful means of conveying details about patient response to therapy not fully delineated by summary statistics, and confirms a valuable treatment benefit of under 3 hour fibrinolytic therapy in acute stroke. (Stroke. 2010; 41:2381-2390.)

Key Words: cerebral infarction • thrombolytic therapy • clinical trial • stroke • acute • charts

Visual representations of the results of clinical trials, when performed correctly, complement and extend numeric statistical analyses. However, improperly performed graphical analyses can mislead and confuse, rather than inform. As Twain might have observed had he met Tufte, in addition to lies, damn lies, and statistics, clinicians and scientists sometimes must contend with lines, damn lines, and graphical representations of data.1 This challenge now faces emergency physicians, neurologists, and other practitioners who care for acute stroke patients.

A faulty figural criticism of the 2 NINDS-tPA trials recently was published,2 departing from best practices in the visual display of quantitative information.1 This latest criticism from antitreatment contrarians does not really threaten to slow adoption and implementation of beneficial fibrinolytic stroke therapy. In the 15 years since publication of the 2 original NINDS-tPA trials,3 the robustness and reliability of their results has been repeatedly confirmed by independent analyses of the trials by specially convened panels, multispecialty societies, and independent regulatory agencies worldwide.4–10 Pooled data from 4 other trials have provided independent evidence of benefit in the under 3 hour window.11–14 Most recently a seventh major trial confirmed that benefit extends to the 4.5 hour time window.15 Regional systems of stroke care which support collaboration between emergency physicians and neurologists are proliferating rapidly throughout the United States and the world to ensure delivery of lytic therapy to an increasing proportion of patients.16,17 Thrombolytic therapy is now an accepted staple of acute stroke care among all cerebrovascular disease experts.

Nonetheless, the recent article has the potential to cause some confusion among nondisease and nonstatistical experts. According to the deviation of the analysis’ faults would be salutary. Moreover, the publication offers a useful occasion to generate new graphical representations of the results of the 2 NINDS-tPA trials that are appropriate to the nature of the data, and can serve as examples of good practice in the visual display of trial findings in acute stroke and in other conditions for which ordinal outcome scales are important study end points.

How to Hide a Treatment Effect

A critical step in the visual display of quantitative information is to space tick marks on graph axes at intervals that are appropriate to the measure being depicted. Consideration
must be given to whether the underlying variable has a continuous, ordinal, or binary distribution, whether there are floor or ceiling effects, and whether scale scores are skewed in their functional significance. Here the recent article by Hoffman and Schriger erred substantially. All 13 of the graphs in the article’s main text map the NIH Stroke Scale (NIHSS) score or absolute change on the NIHSS (delta NIHSS) on one of the axes. The NIHSS and delta NIHSS scores are both ordinal and severely skewed in their functional significance. However, contrary to these well-known aspects of the NIHSS, all of Hoffman and Schriger’s graphical representations depict the NIHSS and delta NIHSS as integer scales of equal functional import over their entire range.

An extreme instance of the NIHSS’s nonlinearity is the nearly total absence in ischemic stroke populations of scores between 32 and 41 on this 42 point scale. Scores in this range are only attainable by comatose patients, and coma is extremely rare at ischemic stroke onset, uncommon in the first 2 weeks after onset, and vanishingly rare as a persistent state at final trial outcome evaluations at 3 to 6 months. The scarcity of NIHSS scores in this range is reflected in the 2 NINDS trials: among the 1248 NIHSS scores obtained in placebo and tPA-treated patients at entry and 3 months, only 7 (0.56%) were between 32 to 41. As a result, this segment of the NIHSS has little informational value regarding outcomes in ischemic stroke populations. The same phenomenon occurs when NIHSS scores at 3 months are subtracted from pretreatment scores to yield an absolute change in NIHSS (delta NIHSS) value. Virtually no patients in standard stroke populations experience delta NIHSS scores in the 22 to 42 or −33 to −42 ranges. In the NINDS-tPA trials, even though these values numerically span 36.5% of the delta NIHSS range, only 2.4% of patient scores fell in these segments. Consequently, Hoffman and Schriger’s approach of graphing the NIHSS from 0 to 42 and the delta NIHSS from −42 to 42 in integer increments of equal visual weight will compress the true possible clinical response range and visually obscure genuine treatment effects. The visual space their approach devotes to the NIHSS is too large by between 35% to 57% (42/31 = 1.35 for NIHSS; 85/54 = 1.57 for change in NIHSS).

A second, critical aspect of the NIHSS’s nonlinearity is the skewed functional significance of scores within the range of values that are attained by patients. Changes in lower numbers on the NIHSS reflect dramatic changes in disability and quality of life, whereas similar changes among higher numbers are of much more minor functional consequence. For example, improving from an NIHSS score of 31 to 21 barely affects patient disability, whereas going from an NIHSS score of 11 to 1 results in a dramatic improvement in daily functioning, although both are a 10 point lowering. Giving lower range score changes equal visual weighting to upper range score changes is misleading.

Because the functional salience of different portions of the NIHSS is unfamiliar to most nondisease experts, it is helpful to illustrate this point with a more well-known example, taken from the everyday experience of all Americans—numeric scores and letter grades on high school tests (Figure 1A to D). Although high school tests are scored using a 0 to 100 point range, the functional significance of these raw scores is highly skewed. Letter grade ranges reflecting important differences in subject mastery are all clustered on the high end of the scale. For example, improving from a test score of 10 to 50 results in no alteration in the letter grade of F, failure. But going from a test score of 55 to 95 results in improvement in the letter grade from F, failure, to A, complete subject mastery. Both are 40 point increases, but one reflects little and the other dramatic improvement in subject expertise. Any graphic display that depicts changes in different portions of the raw test score scale as visually equal is perceptually misleading to the viewer, underemphasizing the 31 point response range (59–90) in which all 4 important grade state transitions occur and overemphasizing the 60 point (0–59) range in which no grade state transition transpires.

An additional graphic problem can arise if raw test score change values (delta test scores) are graphed along an axis that far exceeds the maximum change in scores theoretically achievable by individual students. For example, a student with a preteaching intervention test score of 75 can only possibly improve to 100 with a successful teaching intervention. This increase from C to A performance is a substantial beneficial result. But if the change in raw test scores are crudely graphed as ranging from −100 to 100, this 25 point change will span only 25/200 = 12.5% of the axis scale and appear as a minimal, rather than substantial, benefit.

The functional significance of the NIHSS, like that of raw high school test scores, is highly skewed. Concurrent ascertainment of NIHSS neurological scores and modified Rankin Scale global disability scores demonstrates that 4 of the 6 important disability state transitions occur in the lower one-fifth of the NIHSS range and a fifth occurs at the extreme upper end20 (Figure 2A). A broad expanse of the scale, 56% (raw score values from 18 to 41), is not associated with any major disability state alteration. Hoffman and Schriger do not take the functional nonlinearity of the NIHSS into account when generating their graphic representations. All 13 of the graphs in their main article fail to visually adjust for the skewed functional significance of the NIHSS scale and 5 map delta NIHSS values along axes 2 to 20 times the range achievable by individual patients, severely compressing the apparent treatment effect and artificially giving the appearance of an effect too small to be of clinical importance (Figure 2A to D).

Compounding these issues, many of Hoffman and Schriger’s Figures are cumulative distribution curves, on which incremental, but clinically worthwhile, treatment advances typically appear as visually modest curve separations. Readers familiar with this property of cumulative distribution curves will recognize the clinical relevance of visually small separations, but unwary readers may be vulnerable to Hoffman and Schriger’s misplaced encouragement to be surprised by this aspect of their figure.

An additional defect of all of Hoffman and Schriger’s main article graphs is that they confound the depiction of the effect of treatment on outcome by concurrently charting the effect of other variables on outcome. Many features determine patient outcome after stroke, including baseline stroke severity, age, history of prior disability, serum glucose, and deployed treatments. Of these, baseline stroke severity has the greatest influence. When the goal of a graphical analysis is to depict all or several determinants of stroke outcome, best
practice figure will incorporate several prognostic variables in their layout. However, in treatment decision-making, the salient issue is whether an active therapy will improve final outcome over that which will occur with supportive care alone. Patients and physicians cannot control the presenting stroke severity—that is a given by the time the patient presents to the emergency department. Clinicians can alter final outcome by selecting the appropriate intervention. Consequently, graphs most relevant to treatment decision-making will elide depictions of the influence of confounding variables on the outcome and focus directly on the relationship of treatment choice and outcome.

To demonstrate how extremely Hoffman and Schriger’s approach stacks the deck against a visual indication of treatment benefit, it is helpful to consider how an unattainably perfect treatment would appear on their lead graphical display. Let us imagine an ideal treatment, call it Miraculous Plasminogen Activator (MPA), which makes every single treated patient have a complete recovery and normal outcome. Such a treatment is of course physiologically impossible, because many patients have already suffered substantial stroke injury before therapy can be delivered. However, were it achievable, MPA would be an unprecedented boon to mankind, ridding humanity entirely of the scourge of stroke, the second leading cause of death and a leading cause of adult disability worldwide. And how does this perfect, unattainable cure appear in Hoffman and Schriger’s figural universe? As a minor advance, barely worth pursuing (Figure 3). The area between the curves separating MPA from placebo would occupy only 17% of the graphical space between the x and y axes. More than four-fifths of the graph region is dead space, irrelevant to treatment effects. Clearly, mapping cumulative distribution functions of the absolute delta NIHSS results in graphics that are perceptually misleading.

Generating Clinically Appropriate Graphs of the NINDS-tPA Trials

Given the functional nonlinearity of the NIHSS, how can clinically appropriate graphical displays of acute stroke trial...
results, including those of the 2 NINDS-tPA trials, be generated? Several methods exist that are appropriate to the graphical depiction of scales with ordinal functional values and skewed population distributions, including: (1) charting normalized gain and loss, (2) graphing clinically relevant ordinal ranks, (3) charting prespecified primary dichotomized outcomes, (4) graphing scores converted to a true continuous value scale, and (5) charting joint outcome table analysis. Overall, the intervention has substantial benefits. For every 100 patients treated, tPA improves final outcomes in 32, including complete or near complete recovery in 13, worsens final outcome in 3, and does not alter final outcome in 65. C, Cumulative distribution graph shown by Hoffman and Schriger of the delta NIHSS results in the 2 NINDS tPA trials. Green lines are tPA groups and blue lines placebo groups. Because of the extreme nonlinearity of raw NIHSS score functional importance, the intrinsic characteristics of cumulative distribution displays, and the axis scaling at a range up to 20 times the change score achievable by individual patients, the lytic intervention appears to have little effect. D, Stacked bar chart of final global disability outcomes in the 2 NINDS-tPA trials delineates the benefit of lytic treatment obscured in the cumulative distribution plot.

Normalized Gain and Loss

Normalized change is a statistically appropriate metric to use when analyzing change across nonparametric scales, with different starting values for individual data points. To derive normalized change scores, an observed change is divided by the numerically achievable change available to that individual. As a result, normalized change scores reflect the proportion of the theoretically possible change accomplished by an intervention. Recent salient work in normalized change analysis has occurred in the physics education literature. Faced with the functional skewness in student test scores (Figure 1A), physicist educators recognized that calculating and mapping crude absolute change was misleading, and developed normalized change analysis as a more appropriate numeric and graphical approach.

When applied to NIHSS, the normalized gain measures the fraction of the possible improvement that is experienced by a patient from his/her pretreatment NIHSS to his/her day 90 NIHSS. The normalized gain is given by the formula: Normalized Gain (%) = \(\frac{(\text{NIHSS}_{B} - \text{NIHSS}_{90D})}{\text{NIHSS}_{B}} \times 100\), where NIHSS\(_{B}\) is the NIHSS score at pretreatment baseline and NIHSS\(_{90D}\) is the NIHSS at the 90 day visit.

The normalized loss measures the fraction of the possible decline that is experienced by a patient from his/her pretreatment NIHSS to his/her day 90 NIHSS. The normalized loss is given by the formula: Normalized Loss (%) = \(\frac{(\text{NIHSS}_{B} - \text{NIHSS}_{90D})}{42 - \text{NIHSS}_{B}} \times 100\).

Figure 2. Data from the 2 NINDS-tPA trials similarly demonstrate importance of incorporating variations of functional significance across scale ranges into visual data displays. A, Bar chart showing the nonlinear functional meaningfulness of NIHSS scores. Important transitions in global disability (reflected in modified Rankin Scale levels 0 to 6) are clustered to the far right or far left of the raw score distribution. R6 = dead, R5 = bedridden, needing constant care, R4 = dependent but able to walk, R3 = independent in daily living but occupationally disabled, R1 = neurologically symptoms without disability, R0 = symptom-free. B, Pie chart showing the findings of the NINDS tPA trials for the effect of tPA on final outcome compared with supportive care, using joint outcome table analysis. Overall, the intervention has substantial benefits. For every 100 patients treated, tPA improves final outcomes in 32, including complete or near complete recovery in 13, worsens final outcome in 3, and does not alter final outcome in 65. C, Cumulative distribution graph shown by Hoffman and Schriger of the delta NIHSS results in the 2 NINDS tPA trials. Green lines are tPA groups and blue lines placebo groups. Because of the extreme nonlinearity of raw NIHSS score functional importance, the intrinsic characteristics of cumulative distribution displays, and the axis scaling at a range up to 20 times the change score achievable by individual patients, the lytic intervention appears to have little effect. D, Stacked bar chart of final global disability outcomes in the 2 NINDS-tPA trials delineates the benefit of lytic treatment obscured in the cumulative distribution plot.
By dividing the observed gain or loss by the possible gain or loss, the normalized change score formulas somewhat cancel out the large sections of the NIHSS that are not clinically relevant, while capturing the changes that are important to each individual patient. For example, consider the mapping of a patient with a baseline NIHSS score of 12 who improves to a score of 0. This patient moves from a severely compromised functional state to fully normal. The patient’s absolute delta NIHSS score misleadingly suggests a modest benefit, a gain of only 12 points on an 85 point scale (from 42 to 42). His normalized gain score suggests a major benefit, a gain of 100% indicating complete resolution of deficit.

A closely related metric to normalized change is percent change. Bruno and colleagues, in their important work highlighting the inappropriateness of using the absolute delta NIHSS to analyze the NINDS-tPA trials, introduced the percent change as an improved approach. For improvements between baseline and day 90, percent change and normalized change yield equal values. For worsenings, the values differ. Because it handles gains and losses in a symmetrically meaningful way, normalized change is theoretically the more appropriate measure. However, because most stroke patients improve from baseline to day 90, in practice in acute stroke trials normalized change and percent change analyses frequently yield similar results.

It is important to recognize that, although the normalized change score substantially attenuates the visual distortions produced by the functional nonlinearity of the NIHSS, it does not completely eliminate them. Some changes on the NIHSS that are of only modest clinical relevance will appear more important than clinically warranted (e.g., going from NIHSS of 30 to 41 is only a mild functional worsening but will appear as a substantial normalized loss of 91%). Nonetheless, normalized gain and loss on the NIHSS are generally much more clinically relevant metrics than absolute score changes.

Figure 4 shows a graph of the normalized gain and loss scores observed in the 2 NINDS-tPA trials. As expected, the distribution of change scores is highly bimodal and non-normal. Improvement tended to occur in all stroke survivors and was generally substantial. Worsening occurred primarily in patients who died and was therefore extreme. A substantial benefit of active tPA therapy was noted. Numerically, tPA was associated with a median 66.7% improvement in NIHSS score (interquartile range, 0 to 92.0) whereas placebo was associated with a 50.0% improvement (interquartile range, −7.1 to 80.0). On average, tPA patients recovered two-thirds of the way toward fully normal at 90 days, whereas placebo patients improved only half-way toward fully normal. The associated Wilcoxon rank sum probability value of 0.003 indicates an overwhelming probability that tPA is beneficial. The graph visually delineates the substantial degree of benefit.

Clinically Relevant Ordinal Ranks

An additional approach to handling data that is strongly nonlinear in functional importance is to use cutpoints to divide the scale into ordinal ranks, each of which has substantial clinical import. For example, this approach is the standard one used in educational testing, in which the highly skewed 0 to 100 raw scores on tests are divided by cutpoints into more salient letter grades. Though Hoffman and Schriger make no mention of them, stacked bar charts reflecting clinically relevant ordinal ranks were an important early innovation in graphical analysis and reporting of acute stroke trials. Stacked bar charts are now the most frequently used and widely accepted graphic technique for display of acute
stroke trial findings. Using clinically relevant cutpoints, outcome scales, including the NIHSS, are divided into 3 to 7 clinically relevant ranked levels, displayed with most desirable outcomes on the left and least on the right (Figure 1D and 2D). The stacked bar chart summarizes final outcome distributions efficiently and accurately and permits viewers to rapidly assess the clinical use of interventions. Trellis displays of multiple stacked bar charts for patients with different pretreatment factors permit identification of modifiers of treatment effect. Stacked bar charts of the NIHSS and other outcome scales for the NINDS-tpA trials were presented in the primary publication reporting both trials and severity-adjusted trellis stacked bar charts in a later analysis, and all visually demonstrated substantial treatment benefit.5,25

Choosing the best cutpoints to use in segmenting an imbalanced raw score scale into a clinically useful ordinal scale can be challenging. Optimal cutpoints may differ for different clinical applications. Several alternative cutpoints may be reasonable to use, without a single, clearly best selection. Hoffman and Schriger argue that their approach of avoiding any segmentation of the NIHSS and graphing the entire raw score distribution reduces the chance that bias will be introduced through the selection of one set of cutpoints rather than another. However, with a scale as ill-suited as the untransformed NIHSS to graphical depiction, retaining the entire raw score scale promotes rather than protects against bias, as it will strongly influence against visualization of beneficial or harmful treatment effect. Any reasonable set of cutpoints will yield graphics that are more clinically relevant and informative than leaving the raw scores untransformed.

While conventional stacked bar charts show only final outcomes, clinically relevant ordinal categorization of the NIHSS also allows generation of displays that depict the evolution of each individual patient’s deficit from pretreatment baseline to the 3 month primary outcome visit. Figure 5 shows a matrix column display delineating the starting and final NIHSS rank status of every patient enrolled in the NINDS-tpA trials. The graph demonstrates a substantial benefit of active, lytic therapy.

**Charting Dichotomized Prespecified Primary Outcomes**

In interpreting the results of phase 3 clinical trials, considerations of prespecified versus post hoc analyses apply as fully to figural as to numeric analyses. Pivotal clinical trials are designed to test a single, prespecified mathematical hypothesis that can be expressed numerically or graphically. Graphs depicting the results of this prespecified analysis are the figures of first and utmost relevance in assessing trial findings. Of far lesser relevance are mathematical analyses of prespecified secondary outcomes. Of least relevance are post hoc graphical analyses of nonprespecified hypotheses (which when misused are a form of “figural data dredging”).

Identifying a single, clinically relevant cutpoint on a nonlinear scale as a “win criterion” is an additional simple but effective approach to handling nonlinear scales. This strategy converts each scale into a binary success/no success measure. In the second, confirmatory NINDS-tpA trial (trial 2), the investigators selected as the primary outcome favorable recovery on 4 different stroke outcome scales, segmenting each of the scales at single, clinically relevant cutpoints to define favorable and unfavorable recovery. Figure 6 shows the graphical depiction of this analysis. This graph is the only figure that is authoritative with regards to whether the trial was positive or negative, and it demonstrates consistent evidence of benefit across all components of the composite end point.

**Converting Scores to a True Continuous Scale**

The ultimate solution to the challenges inherent in graphing ordinal scales and scales with nonlinear functional distributions is to avoid the problem altogether, by graphing a scale that is continuous and that has transitions of equal clinical import over its entire range. Recent work in health outcomes research has made this approach feasible in the analysis of acute stroke clinical trials.

A variety of mathematical techniques, including standard gamble, time-tradeoff, person tradeoff, and visual analog scales enable the assignment to diverse diseases and disease intensities a continuous score indexing the resulting degree of deviation from optimum health. Tasked with assigning valu-
ations to all human diseases, the World Health Organization Global Burden of Disease Project (WHO-GBDP) used the person trade-off method to generate disability weight (DW) values for each condition, using 22 conditions distributed over the outcome range as anchors around which to place additional health states. Using the WHO-GBDP person trade-off method, a disability weight value can be derived for any health outcome state. DWs range from 0 (normal health without disability) to 1.0 (dead or as bad as being dead).

A recent study used the WHO methodology to derive DWs for each of the 7 modified Rankin Scale ordinal levels of disability.26 As a result, the ordinal, functionally skewed levels of the mRS can be converted to continuous, functionally evenly spaced DWs. Related work has also shown that NIHSS values can be mapped to the modified Rankin Scale;27 consequently, NIHSS levels can also be converted to DWs. Figure 7 shows the disability weight outcomes for every patient in the 2 NINDS-tPA trials, mapping their evolution from before treatment assignment to the final follow-up year after stroke. The Figure demonstrates a marked and sustained benefit of therapy in reducing the degree of disability among treated patients throughout the year following therapy.

Joint Outcome Analysis Charts
Joint outcome analysis provides clinicians and patients with the most explicit understanding of the benefits and risks of a therapy. For a group of 100 or 1000 patients, joint outcome tables numerically show the outcomes the cohort will achieve if they receive control therapy and the outcomes they will achieve if they receive active intervention.27,28 Similarly, graphical depictions of joint outcome tables provide the clearest figurative delineations of a therapy’s impact. For crossover design trials, joint outcome tables can be computed directly from trial data results. For parallel group design trials, joint outcome tables can be computed by algorithmic, repeated simulation, and expert panel derivation techniques constrained by trial data. Figure 8 shows charts of the joint outcome results of the 2 NINDS-tPA trials. The figures again demonstrate substantial net beneficial effects.
Concluding Comments: Looking Backwards and Going Forwards

This detailed analysis has revealed substantive flaws in all the main article figures presented by Hoffman and Schriger. Every one of their graphics fails to visually adjust for the extreme nonlinearity of the NIHSS.

In contrast, we have presented 12 different graphs of the 2 NINDS-tPA trials that avoid the incorrect handling of the NIHSS present in Hoffman and Schriger’s figures. All display a substantial benefit of early fibrinolytic therapy. Three of these figures depict the starting point and final outcomes of all patients enrolled in the trials (Figures 4, 5, 7), an avowed goal of Hoffman and Schriger, but unlike their figures, do so taking a visually appropriate approach to the nonlinear properties of the NIHSS. We have also generated a range of figures that display outcomes in summary, rather than all data, form, an additional important aspect of graphical analysis. We have presented a wide array of figure types, including: bar charts of normalized gain and loss; stacked bar, bar, and matrix plots of clinically relevant ordinal ranks; a time series stacked line plot of continuous scale disability weights; and line plot, bubble chart, and person icon array graphs of joint outcome table analysis. Several of these figural techniques are new in acute stroke trial analysis and may serve as useful formats for presentation of results of future trials and starting points for further innovations in figural presentation. All visually demonstrate substantial benefit of therapy.

Given the robust graphical and statistical evidence of benefit seen in the 2 NINDS-tPA trials, and supporting data

Figure 7. Stacked area plot of disability burden throughout the first year after stroke among all patients enrolled in the 2 NINDS-tPA trials. Color bar shows the hue intensities assigned to the poles and midpoint of the WHO disability weight, and displayed colors reflect the exact value of the DW along this continuous spectrum. Patients in both treatment groups begin with severe disability burdens (red and orange predominate at 0 timepoint). Among those who survive their stroke (numerically more with tPA than placebo throughout the 1 year period), substantially less disability is experienced by tPA patients throughout the first poststroke year.

Figure 8. Joint outcome figures based on data from the 2 NINDS-tPA Trials. A, Line plot showing final global disability outcomes of 100 patients if they receive placebo or tPA treatment. Green indicates better outcome with tPA, red worse. B, Bubble plot of same data. C, Person icon array, decision-aid figure showing among 100 treated patients the change in outcomes that will occur with tPA rather than placebo. (Figure 8C reproduced with permission of UCLA Stroke Center).
for benefit from under 4.5 hour patients enrolled in 5 other tPA trials, the question arises: whence does the resistance to this therapy among a residual segment of contrarians arise? When thrombolytic stroke therapy was introduced in the 1990s, it was a disruptive medical innovation that could not easily be incorporated into existing patterns of care. Neurorlogists resisted the changes in practice that lytic therapy required, including the need to respond emergently to the emergency department.\(^{29}\) Radiologists resisted reading CT scans emergently, rather than the following morning during routine reading sessions.\(^{30}\) Emergency physicians resisted having to give a therapy with risks as well as benefits, for a complex brain disease, unsupported by neurologist, radiologist, and institutional backing.\(^{31}\)

Consequently, the evolution in care that the advent of lytic stroke therapy constituted evoked a conservative counter-reaction. One form this reaction took was to deny the therapy was beneficial at all. If the treatment did not confer benefit, familiar patterns of practice would not need to be changed. Denying that thrombolytic stroke therapy is beneficial is no longer (indeed never was) tenable, and yet it persists, albeit with much less influence than in the past, slowing the adoption of thrombolysis and the buildout of collaborative regional stroke systems of care.

It is important to be clear-eyed about thrombolytic therapy for stroke. It is far from a miraculous cure.\(^{27}\) Only about 32 of every 100 treated patients benefit in a clinically important manner—we would like that to be 100 of 100. Only about 13 of every 100 treated patients end up normal or near normal as a result of therapy—we would like that to be 100 of 100. As many as 3 of every 100 treated patients has a worse final outcome as a result of treatment—we would like that to be 0 of 100. The 2 NINDS-tPA trials were only the start of the therapeutic era for acute ischemic stroke. Better therapies are urgently needed and actively being developed.

But intravenous thrombolysis with tPA does work to a substantial degree. The development of thrombolytic stroke therapy was one of the signal triumphs of emergency medicine and neurology at the turn of the twentieth century. Emergency medicine and neurology physicians, collaborating with hematologists, neurosurgeons, radiologists, nurses, pharmacists, statisticians, and allied health personnel, designed and performed the pivotal trials demonstrating benefit and then designed and implemented regional systems of stroke care to permit rapid, reliable, and safe delivery of therapy to patients in routine practice. As the analyses presented in this article demonstrate, emergency physicians and neurologists can be assured that intravenous thrombolytic therapy conveys substantial benefit, and is strongly supported by the entire array of modern clinical trial analytic techniques, both statistical and graphical.

**Sources of Funding**

J.L.S. and S.S. are funded by NIH-NINDS Awards P50 NS044378 and U01 NS 44364, and an American Heart Association PRT Health Outcomes Research Center Award.

**Disclosures**

J.L.S. is a scientific consultant regarding trial design and conduct to CoAxia, Concentric Medical, Talecris, Ferrer, BrainGate, PhotoThera, and Sygnis (all modest); has received lecture honoraria from Ferrer (modest); has declined consulting/honoraria monies from Genentech since 2002; is a site investigator in the NIH MR RESCUE, CLEAR-ER and IMS 3 multicenter clinical trials for which the UC Regents receive payments based on the clinical trial contracts for the number of subjects enrolled; has served as an unpaid site investigator in multicenter trials run by Lundbeck and NTI for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; is a site investigator in a multicenter registry run by Concentric for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; is an employee of the University of California, which holds a patent on retriever devices for stroke. J.G. is an employee of the University of California, which holds a patent on retriever devices for stroke. S.S. is a site investigator in the NIH MR RESCUE, CLEAR-ER and IMS 3 multicenter clinical trials for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; has served as a site investigator in a multicenter trials run by Lundbeck and NTI for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; is a site investigator in a multicenter registry run by Concentric for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; has served as a medicolegal expert defender of emergency physicians in acute stroke tPA litigation; administers stroke thrombolytic therapy in his practice (less than 5% of effort); and is an employee of the University of California, which holds a patent on retriever devices for stroke.

**References**

Graphic Reanalysis of the Two NINDS-tPA Trials Confirms Substantial Treatment Benefit
Jeffrey L. Saver, Jeffrey Gornbein and Sidney Starkman

Stroke. 2010;41:2381-2390; originally published online September 9, 2010;
doi: 10.1161/STROKEAHA.110.583807
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2010 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/41/10/2381

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/